arXiv:0705.1130v2 [hep-th] 9 Feb 2009

An Entropy-Weighted Sum over Non-Perturbative Vacua

Andrea Gregori[†]

Abstract

We discuss how, in a Universe restricted to the causal region connected to the observer, General Relativity implies the quantum nature of physical phenomena and directly leads to a string theory scenario, whose dynamics is ruled by a functional that weights all configurations according to their entropy. The most favoured configurations are those of minimal entropy. Along this class of vacua a four-dimensional space-time is automatically selected; when, at large volume, a description of space-time in terms of classical geometry can be recovered, the entropy-weighted sum reduces to the ordinary Feynman's path integral. What arises is a highly predictive scenario, phenomenologically compatible with the experimental observations and measurements, in which everything is determined in terms of the fundamental constants and the age of the Universe, with no room for freely-adjustable parameters. We discuss how this leads to the known spectrum of particles and interactions. Besides the computation of masses and couplings, CKM matrix elements, cosmological constant, expansion parameters of the Universe etc..., all resulting, within the degree of the approximation we used, in agreement with the experimental observations, we also discuss how this scenario passes the tests provided by cosmology and the constraints imposed by the physics of the primordial Universe.

[†]e-mail: agregori@libero.it

Contents

1 Introduction

In this work we discuss the physical scenario arising from the condition finiteness and universality of the speed of light are paired with the related hypothesis of existence of a horizon to our observations, when the space enclosed within the horizon is viewed as the *whole* effectively existing space [1]. Everything within such a space is causally connected, through light rays, to the observer, for which the horizon surface turns out to correspond to the origin of the Universe, intended as the whole of space-time and the physical content effectively accessible through experiments (from this respect, the Universe can be equivalently defined as the region causally related to us, the observers).

We will find that these conditions necessarily lead to a quantum, string theory (or "M-theory", if one prefers) scenario, which precisely corresponds to our Universe, with the known interactions and particles/matter content. Quantum mechanics itself turns out to be not an independent, additional input, but a consequence of these two starting conditions. The physics of the system is ruled by a functional, 2.31, which weights all configurations according to their entropy. There is no "solution" for "the" configuration of the Universe in a classical sense, but only a "mean value", which approaches the better and better a kind of classical limit at large space-time volumes, where some aspects of the physical world admit an approximated description in terms of ordinary cosmology and relativistic quantum

[1]Notice the difference between this point of view, and the usual interpretation of the space within the horizon of observation as only the *portion* of the Univers accessible to our observation. In the usual case, we have a truncation, or restriction, of a possibly wider space, to a subregion, which inherits the geometry from the wider space; in our case, we have an *absolute* space, that, owing to the fact that the surface at the horizon topologically corresponds to a neighbour of the point at the origin of the universe, turns out to be necessarily curved.

field theory. In our concluding remarks, section 11, we comment on the hypothesis that this functional is of even more general validity, automatically selecting the configuration of the Universe corresponding to our experience out of a general class of configurations, well beyond critical string theory.

Indeed, since the time of their appearance it is a long debated question, whether General Relativity and Quantum Mechanics can be accommodated in a "unified" conceptual description. The seek for such a theoretical framework finds its major difficulty in the non-renormalizability of gravity, when intended as a theory of fields, such as electromagnetism is. Along the time, String Theory arose as a good candidate, in that it consists of a theory of objects with a non-trivial geometry and a built-in quantizable harmonic-oscillator structure, such as is required in order to describe excitations corresponding to fields and elementary particles. The investigation in this direction received a big impulse when it was realized that quantum strings are renormalizable, and even more when, more recently, strong evidence has been produced that String Theory is unique: according to this idea, there should be a unique theoretical structure underlying all string constructions; these would then be "slices" of a unique theory, covering certain regions and corresponding to certain limits in the moduli space. What this theory precisely is, it is not yet clear; whether M-theory, conformal or in general not conformal etc... Whatever this general theory underlying all perturbative slices is, we will refer to it simply as to "String Theory".

A common approach to string theory is to consider it as a "source" for terms of an effective action. In most cases, they are derived with geometrical methods based on differential and algebraic geometry. These entries are then treated by making large use of field theory techniques. The non-perturbative properties of a "string vacuum" are inferred through an extensive use of string-string duality. In any case, the approach to string theory is somewhat "hybrid", strongly anchored to a way of seeing inspired by the geometric (general relativistic) and field theoretic (quantum field) points of view. However, quantization of gravity basically implies quantization of space-time itself, and this somehow "destroys" geometry in the traditional sense. At least, for sure it destroys differential geometry. A continuous, differentiable space-time arises from quantized space-time coordinates only as a large distance limit, and in general only as an approximation, valid under specific conditions: the limiting procedure is not so straightforward and conceptually simple, in a theory basically characterised by a built-in T-duality. For instance, if T-duality is not broken, in general there can be singularities which are not smoothed down by going to a large volume limit.

On the other hand, it remains unexplained why should space-time be quantised, i.e. why should we look for a quantised version of gravity, apart from the evident analogy of the gravitational force, in the weak coupling regime of the Einstein's equations, with the wave equations of electromagnetism (and, even before it, the Coulomb form of the interaction common to both the gravitational and the electric force).

On the theoretical side, some attempts have been made, of describing quantum mechanics in terms of objects proper to classical mechanics, through a statistical description of quantum amplitudes. After all, quantum mechanics is a kind of "modified classical mechanics", in which the Poisson brackets are substituted by (anti)commutation rules, and the theoretical framework of quantum mechanics makes extensive use of technical tools and theoretical

concepts of statistical mechanics, such as "mean values", applied to harmonic oscillators: "wave amplitudes", "wave decompositions/wave superpositions" etc... In this work, we start by considering from the very beginning the implications of the ideas of General Relativity and the finiteness of the speed of light. The fact that light travels with a constant, finite speed, not only has the well known consequences for what concerns the way we relate observations made in different frames, moving with respect to each other. It also implies a deep modification of the "geometry" of space-time, and is basically at the origin of the curvature of the Universe, if this is intended as the region causally connected to the observer and, therefore, the region indeed accessible to experimental observations. We will discuss how the curvature of the bounded space is basically due to the fact that the light rays that appear to us as coming from a horizon surface surrounding us with a full, 4π solid angle, indeed originate from one "point" (the Big Bang point). Under these conditions Special Relativity implies in fact that a space-time, a "Universe", with a certain age, possesses a non-vanishing curvature, which, according to the equations of General Relativity, can be seen to correspond to an energy. An inspection of this quantity reveals that this ground energy precisely corresponds to the minimal energy fluctuation we would expect for an object extended as the Universe, according to the Heisenberg's Uncertainty Principle. The uncertainty relations which are at the base of quantum mechanics appear therefore to be naturally embedded in a Universe governed by the laws of General Relativity. At this point one may wonder whether Quantum Mechanics is just a convenient parametrization of a world whose oscillatory, wave-like appearance, is the consequence of the fact that all what we can observe comes to us mediated by either light or gravity, in any case through a set of waves that propagate at finite speed c.

Further inspection of General Relativity in the framework of a "universe" with finite horizon reveals that the underlying description, besides a quantum nature, must also possess a T-duality symmetry. This seems to select String Theory for a description of such a Universe. In this framework, many old questions are addressed in a completely different way. Among the features of space-time is the existence of a minimal length, such that it does not make anymore sense to talk about "open" or "closed" geometric sets in the traditional sense: the "dimensionless" point does not exist. Its substitute, the Planck-size cell, is then "topologically equivalent" to a disk, or to a ball as well. The equivalence of the horizon surface to the Big Bang point can be understood only in this new topology. Classical geometry is only an approximation, valid at large space-time volumes.

Under the hypothesis of uniqueness of string theory, for which strong evidence, although not yet a real proof, has been produced, we arrive to our proposal for a functional encoding all the information about the evolution of the "Universe", i.e. the dynamics of space-time and the matter it contains, expression 2.31 [2]. This functional is derived from simple first principles, no particular dynamic information being introduced as external input, and consists of a sum over all string configurations, weighted by a function of their entropy. The dynamics comes out as a consequence of the fact that the system "evolves" with higher probability from a certain configuration to closer, "neighbouring" configurations which occupy a higher volume in the phase space, therefore preferably through steps of minimal increase of entropy.

[2] In Ref. [1] an equivalent expression is derived in a more general framework, not relying on string theory. We discuss there what is the role played by string theory in this more general context.

At any step, what we experimentally observe is a superposition of configurations, in which those with the lower entropy dominate, and what we interpret as "time" evolution is indeed an ordering along the target space volumes of the class of dominant configurations. At any "time", these correspond to the maximal possible breaking of symmetry. The dominancy of the contribution of these configurations to the mean value of any observable increases the more and more "as time goes by", and the Universe cools down. We have in this way a realization, at a non-field theory level, of the idea of spontaneous breaking of symmetry, because the mean value of observables effectively shows a "progress" toward more broken configurations. As we will discuss, in these configurations only a four-dimensional subspace is allowed to expand, with the speed of light, and indeed, owing to the presence in the spectrum of massless fields, it expands. Along these four coordinates, T-duality is broken. At large volumes a time-like coordinates can therefore be identified with what we ordinary call "time". In the "classical" limit, i.e. in the class of dominant configurations at large time, the functional 2.31 can be shown to reduce to the ordinary path integral. What we propose appears therefore as the natural extension to quantum gravity of quantum field theory. The concept of "weighted sum over all paths" is here substituted by a weighted sum over all string configurations. The traditional question about "how to find the right string vacuum" is here surpassed in a way that looks very natural for a quantum scenario: the concept of "right solution" is a classical concept, as is the idea of "trajectory", compared to the path integral. The physical configuration takes all the possibilities into account. As much as the usual path integral contains all the quantum corrections to a classical trajectory, similarly here in the functional 2.31 the sum over all configurations accounts for the corrections to the classical, geometric vacuum.

The next step is to test this idea; for this, we must extract from the functional 2.31 physical informations, and compare them with experimental data. The sum 2.31 implies that, at any volume, the physical configuration is mostly (i.e. "looks mostly like") the one of minimal entropy. In first approximation, our analysis can be reduced to see what is the physical content of such a string vacuum.

So simple is expression 2.31, so complicated is obtaining an explicit solution! The point is that the "perturbation" is in this case performed around a non-perturbative string configuration. In order to "solve" the theory, we must use any information coming from non-perturbative string-string dualities. We will see how a good class of representative non-perturbative string configurations, to serve as starting point of our approximation, is constituted by the Z_2 orbifolds. We first investigate entropy in this class of vacua, and, with an extensive use of string-string duality, we find out what is the "spectrum" of the minimal entropy configuration. Fortunately, for this step of the analysis, much of the technology can be borrowed from ordinary results of string theory. In a second time, we move away from the orbifold point, toward the "true" minimal entropy configuration. This process requires to switch on some of the moduli that were frozen at the orbifold point.

It turns out that, in the class of configurations which dominate in 2.31, four-dimensional space-time is automatically selected, and supersymmetry is broken at the Planck scale. The fact of considering space-time as always of finite extension, bounded by the horizon corresponding to a light distance equivalent to the age of the Universe, implies a deep

change of perspective in the computation of several quantities. The reason is that space-time translations are no more a symmetry of the system, but constitute rather an evolution of it. As a consequence, string expansion coefficients can no more be subjected, as usual, to a finite-volume normalization, obtained by dividing them by a space-time volume factor, i.e. by the volume of the group of translations. String amplitudes correspond now to global quantities, not to densities. For instance, owing to the non-existence of low-energy supersymmetry, in this scenario the "vacuum energy" turns out to be of order one in Planck units. Nevertheless, once pulled back into an effective action, it results in the correct value of the cosmological constant. In order to obtain this parameter, we must in fact divide the result of the string computation by the appropriate Jacobian of the coordinate transformation from the string to the Einstein's frame; this introduces a suppression corresponding to the square of the radius of space-time. In this scenario, the "vacuum energy density" is therefore not a constant, but scales as the square of the inverse of the age of the Universe.

A second, important consequence of the missing space-time translational invariance is that the energy of the Universe in not anymore conserved. The energy density of the Universe can be seen to scale, like the cosmological constant, as the inverse square of its age. Indeed, an almost exact symmetry of the dominant string configurations predicts not only a scaling, but also a value of the present-time energy, matter and cosmological densities, of the same order of magnitude. A scaling of these quantities like the inverse square of the age of the Universe implies that the total energy of the Universe scales as its radius. The behaviour, and the resulting normalization, of the total energy, allow to see the Universe as a Black Hole. This point of view is supported also by the computation of entropy, which turns out to follow an area law, scaling with the surface of the horizon, as expected in a black hole.

Inserting the values of the three energy densities in a FRW Ansatz for the Universe, we can then solve the equations and obtain the geometry (i.e. the large scale geometry) of space-time. As it could have been argued from the fact that the horizon is "stretched" from the expanding light rays, the expansion of the Universe turns out to be not accelerated. Nevertheless, to our observations it appears to be: what we observe is in fact a time-dependent red-shift effect, whose time variation, that could be interpreted as due to an accelerated expansion, is produced by a time-variation of the energy and matter scales.

The low-energy spectrum turns out to correspond to the known set of elementary particles and fields. Besides the absence of low energy (i.e. sub-Planckian) supersymmetry, this scenario is characterised by the non-existence of a Higgs field: masses are explained in a different way, and their origin is somehow related to the breaking of space-time parities. They are produced by shifts along the space-time coordinates, that lift the ground energy of a particle. These shifts have a non-trivial effect, because space-time is compact. From a field-theoretical point of view this would lead to inconsistencies of the theory, that would loose its renormalizability. In our framework, however, field theory in an infinitely extended space-time is only an approximation. The right framework in which to look at these phenomena is string theory in a compact space-time; in such a framework, masses can be consistently generated without Higgs fields. The chiral nature of weak interactions is also a consequence of these parity-breaking shifts. The separation of the matter world into weakly and strongly coupled is instead a consequence of the breaking of a T-duality along one internal coordinate.

From a low-energy effective point of view, this appears as an S-duality. Indeed, the shifts that give rise to masses breaks not only parity and time reversal, but also explicitly the group of space rotations. This agrees on the other hand with our experience of everyday in the macroscopical world: the distribution of localizable (i.e. massive) objects in the space breaks in some way the absolute invariance under a change in the direction of observation. Indeed, the functional 2.31 describes the Universe "on shell", and, owing to the embedding of entropy in the fundamental description of physical phenomena, all the symmetries which appear to be broken at a macroscopical level are broken also in the fundamental description. The two levels are therefore sewed together without conceptual separation.

In the class of string vacua dominating our physical world, namely, the minimal entropy configurations, matter is basically non-perturbative: the coupling of the matter sector is one, in Planck units. From this ground value depart the electro-magnetic, weak and strong coupling; the first two running toward lower values, as the space-time volume increases, the third one toward higher values. This poses a fundamental problem to the investigation of the matter degrees of freedom, due to the fact that there are particles which feel both strong and weak interactions. An explicit, perturbative representation of particles as elementary states can only be realized through an expansion around a vanishing ground coupling. Since couplings unify only at the Planck scale, i.e. when we can no more speak of "low energy world", as a matter of fact there is no scale at which all the matter degrees of freedom appear all at the same time as perturbative. If we want to see all matter degrees of freedom explicitly represented in a perturbative construction, such as they appear in usual field theory models, with leptons and quarks all present as ingredients of a low-energy spectrum, we must go to a picture in which the internal coordinates are "decompactified", without moving the internal moduli. In general, owing to T-duality, through decompactification we can have access only to a part of the full theory. What we need is therefore not a true limit: it rather corresponds to a logarithmic representation of the string coordinates.

Indeed, any perturbative string orbifold construction corresponds to a linearized representation of the string space: it is in fact built as an expansion around the zero value of a coupling, and, from a non-perturbative point of view, the latter is a coordinate of the internal space. A perturbative construction implies therefore always a "decompactification" of at least part of the space. This process is non-singular and preserves all the properties of the physical vacuum only if the space under question is flat. Otherwise, as in the cases of interest for us, i.e. of configurations with the maximal amount of orbifold twisting, it is only an approximation, corresponding to considering just the tangent space around a certain point. This reflects in the fact that the contributions of the various coordinates in the computation of mean values appear to be summed, instead of multiplied. A consequence of this artificial "linearization" of the physical space is that couplings appear to run logarithmically with the cut-off mass scale [3]. In a logarithmic representation of space, the vanishing of the ground coupling implies also the vanishing of the tuning parameter of the supersymmetry breaking. As a result, the linearized, perturbative representation, in which all particles show

[3] In a pure flat configuration, such as the vacuum with the highest amount of supersymmetry and no projections at all, mean values such as the mean vacuum energy, or the renormalization of couplings, indeed vanish.

up as elementary states, appears to be supersymmetric, as is the case of many perturbative approaches to string and field theory. On the other hand, in the real world there is no regime in which both leptons and quarks appear at the same time as elementary, weakly coupled, free, asymptotic states.

An artificial linearization of space-time is the cause of another false appearance of the string vacuum, namely the fact that from some respects string theory seems to require for its complete description more than 11 coordinates. Indeed, the 12-th coordinate should be better viewed as a curvature. As is known, we can represent an n-dimensional curved space in n dimensions, with an "intrinsic" curvature, or we can embed it in a $n+1$-dimensional flat space. The degrees of freedom are in any case the same, because in the second case we don't consider the full $n+1$-dimensional space, but an n-dimensional sub-manifold. A perturbative representation of M-theory is something of this kind: when the vacuum corresponds to a curved space, by patching dual representations we have the impression that more than eleven coordinates are required in order to describe the full content, because these representations are necessarily perturbative and therefore built on a flattened, "tangent space".

Owing to their intrinsically non-perturbative nature, investigating the masses of elementary particles, and their couplings, is somehow a "dirty" job, as compared to the elegance of an expression like 2.31. To this purpose, the ordinary machinery of perturbative expansion in Feynman diagrams is of no help, because it applies to a "logarithmic" representation of the real physical world we are interested in. Differential geometry is not appropriate for this deeply non-perturbative string world, and field theory tools can only help in getting some partial results in an approximated, unphysical regime: new tools must therefore be used. For the computation of masses and couplings, we make therefore use of what could be called a "thermodynamical" approach. The idea is that, since, according to 2.31, the entire dynamics of the system is encoded in its entropy, and couplings and masses determine the interaction and decay probability of particles and fields, masses and couplings must be related to the volume occupied in the phase space by the corresponding matter and field degrees of freedom. The problem of computing these parameters in then translated to the one of computing the fraction of phase space these particles and interactions correspond to; this will allow us to determine the "bare" mass and coupling values, i.e. the parameters which are usually considered as external inputs in any effective action. Our approach is therefore somehow reversed with respect to the traditional one. Usually, the parameters and terms of the effective action are used in order to compute the full bunch of interactions. From a general point of view, these can be seen as "paths" coming out from, or leading to, a particle, or in general a physical state. Their amount and strength can be considered therefore a measure of the entropy of the state: the higher is the mass of a particle, the higher is its interaction/decay probability, because higher is the number of final states it can decay to. Entropy is then computed as a function of the interaction/decay probability, in turn determined by the dynamics. In our approach, things go the other way around: it is the dynamics which, consistently with 2.31, is viewed as being determined by entropy. For some respects, this approach can be considered a kind of "lift up" to the string level of those

based on the computation of masses and couplings out of the volume of their phase spaces.

In our framework, couplings and masses turn out to scale as powers of the inverse age of the Universe. They naturally unify at the Planck scale. There is no much to be surprised for the fact that, in usual field theory models, supersymmetry seems to improve the scaling behaviour of couplings, making possible their unification. If they correspond to a "logarithmic" representation of the physical vacuum, couplings unify because, roughly speaking, they are logarithms of functions that unify. In our framework, a logarithmic scaling appears if we want to compare the "bare" values we obtain, with the parameters of a low-energy effective action, in which space-time is considered as infinitely extended. This step in necessary if we want to make contact with the literature. It happens in fact quite often that data of experimental observations are given as result of elaborations carried out within a certain type of theoretical scheme. This in particular is the case of effective couplings and masses run to the typical scale of a physical process. In this case, passing from a large but anyway finite space-time volume to an infinitely extended one results in a "mild", logarithmic correction to the "bare" mass or coupling. Logarithmic corrections work in this case for small displacements in the "tangent space". The bare parameters are instead derived in the full space, and their running is exponential with respect to the one on the tangent space.

In general, any contact between our computations and the data found in the literature must be established at the level of experimental observations, rather than on effective action parameters, whose derivation always depends on a specific theoretical scheme. Therefore, to be rigorous, better than effective couplings one should directly consider scattering amplitudes and decay ratios; one should explain the emitted frequency spectra rather than trying to match given acceleration parameters of galaxies, and so on. This requires a deep change of perspective and a thorough re-examination of any known result. On the other hand, in all the cases a prejudice-free re-analysis of already known results is carried out, we find that our theoretical framework provides us with a consistent scheme. Although almost any physically observable quantity receives a different explanation than in traditional field theory or cosmology approaches, it is nevertheless consistent with what experimentally measured. Indeed, precisely the high predictive power of this theoretical scenario, due to the fact that there are no free parameters that can be adjusted in order to fit data, enhances the strength of any matching with experimental results: any discrepancy could in fact rule out the entire construction. Because of this, a large part of our investigation has been devoted to re-analysing the most important data and constraints, coming not only from elementary particles physics but also from astrophysics and cosmology. Our predictions and results are compatible with any experimental datum we have considered, within the degree of approximation introduced in our derivation.

In this scenario, there is no "new-physics" below the Planck scale. This does not mean that no stringent tests can come from future high energy experiments: for instance, neutrinos turn out to be massive, and what is now a pure prediction could result in a near future into a constraint. However, in practice no major breakthrough is expected from high energy particle colliders, apart from a refinement in the measurement of parameters of already known interactions and particles. However much deceiving this may be (at least from a certain point of view), in this scenario everything is explained within the known matter

states and interactions, although in a new theoretical framework, that shares with field theory and the usual geometrical approach to string theory only some technical similarities.

This work improves the analysis and corrects the results of [2], where the general idea was first presented, although in a incomplete form. For instance, at the time of writing [2] we thought that assuming General Relativity and compactness of the full space-time was not sufficient to fix all the properties of the Universe: Quantum Mechanics was regarded as an independent input, to be added to General Relativity in order to complete the specification of the nature of physical phenomena. Now we see the Uncertainty Principle itself as a consequence of the existence of a horizon to our observations, in a Universe governed by the laws of General Relativity. Our analysis started in Ref. [2] by making the hypothesis that the underlying theory governing the Universe is String Theory. We realize now that also this assumption was redundant. But the major point is that now we have been able to give a formal expression to the idea of superposition of string vacua, weighted according to their entropy. Several statements concerning the minimal entropy configuration have then been corrected: having at hand a deeper understanding of the behaviour of masses and the geometry of space-time, we revised our considerations about the expanding Universe, concluding that the acceleration is only apparent. At the time of [2] also the scaling of couplings was not known, and, inspired by traditional field theory, we supposed it to be logarithmic: our approach suffered from being still too much anchored to ideas belonging to field theory. Therefore, we have now revised also the content of Ref. [3], about the effects of a time variation of the fine structure and masses on atomic energy spectra.

1.1 The outline of the work

The work is organised as follows. We start in section 2.1 by investigating the consequences of the finiteness of the speed of light in a Universe limited to causal region connected to the observer. We see how the geometry of a sphere is implied by the fact that the horizon corresponds to the Big Bang point; our causal region results to be equivalent to a space in which the horizon surface is a "point". In 2.2 we discuss then how the Heisenberg's Uncertainty Principle is naturally embedded in this scenario, and how the same considerations, applied to the case of a particle, lead to the usual uncertainty relations between time and energy, space and momentum. The natural implementation of these properties is therefore a quantum mechanics scenario. Once realized that the Universe shows a wave-like nature and its configurations must be governed by the laws of probability, it is a little step to arrive, in section 2.3, to expression 2.31, a functional encoding all the "dynamics" of this quantum gravity system, which turns out to be a "superposition" of string configurations. We discuss there also how, in the "classical" limit, this functional reduces to the ordinary path integral, and how the exponential weight reduces to the ordinary exponential of the action.

Expression 2.31, the achievement of section 2, can also be considered the starting point of any further analysis, and could be given as "initial input" of a theoretical framework. Therefore, at the end of section 2.3 we pause and list the results in a brief summary. All the following part of the work is devoted to extracting from 2.31 informations about the physical configuration of the Universe.

In the subsequent section 3 we address the problem of how to concretely compute entropy in string configurations. Fortunately, more than an absolute determination, what matters for our purposes is finding out the configurations in which it is minimized. We approach the solution by investigating orbifolds, a choice that we also justify. In section 3.2 we discuss then, within this class of configurations, how the breaking of supersymmetry takes place (subsection 3.2.1), and how a four-dimensional space-time is automatically selected (subsection 3.2.2). We discuss then the origin of masses and the observable spectrum of the theory. We conclude with a discussion about the fate of the Higgs field, not present (and not needed) in this scenario (subsection 3.3.1), and a comment on the breaking of the Lorentz invariance, in particular of the subgroup of space rotations (subsection 3.4), which can explain the observed slight inhomogeneities of the Universe when observed in different directions.

Once identified, with help of the approximation via orbifold constructions, the configuration of minimal entropy and the low-energy spectrum, the next task is to compute observables out of its degrees of freedom. Section 4 is devoted to a discussion of the relation between string amplitudes and effective action parameters. We consider the meaning of the string partition function and mean values within a context of compact space-time and broken supersymmetry. In particular, the condition of broken supersymmetry is identified as the "normal" vacuum configuration, in which, owing to the non-vanishing of the vacuum energy, string amplitudes can be unambiguously normalized. Further implications of being string theory defined on a compact space, and its missing invariance under the group of space-time translations, with the consequent interpretation of string amplitudes as densities, are also discussed.

Once the set up is clarified, we are in a position to compute observables. In section 5 we determine the energy density of the Universe, i.e. the cosmological constant and the matter and radiative energies. We also discuss how, as a consequence of an underlying symmetry of the dominant string configuration, a $''S'' \leftrightarrow$ "T'' $\leftrightarrow$ "U'' symmetry among the sectors corresponding to gravity, matter and radiation, these three contributions to the curvature of space-time turns out to be basically equivalent. The equivalence is not absolutely exact, because also the $''S'' \leftrightarrow$ "T'' $\leftrightarrow$ "U'' symmetry is eventually broken by entropy minimization. However, the breaking is "soft", and leads to a difference between these quantities of the second order. Once these quantities are known, by inserting them in a Robertson-Walker Ansatz for the metric of space-time we can directly verify that the resulting geometry is at the first order the one of a 3-sphere, as it was proposed in section 2.1, on the basis of an analysis of the paths of light rays and the way space-time "builds up" as time goes by. Being this the dominant configuration in 2.31 at large volumes, we conclude that the Universe is, at large times, well approximated by a "classical", FRW description. We discuss here also how the total energy content of the bounded space-time, as well as the total entropy of the Universe, allow to identify it with a black hole.

In section 6 we pass then to the determination of mass scales. At first we consider (subsection 6.1.1) a quantity that can be non-perturbatively computed in an exact way: the "mean mass" in the Universe, namely the eigenvalue of the Hamiltonian at any finite space-time volume. This scale can be seen to basically correspond to the mass of stable

matter: if the matter present in the Universe was constituted by particles all of the same kind, these would have a mass precisely corresponding to this scale. In practice, it roughly corresponds to the neutron mass. This observation somehow agrees with the interpretation of the Universe as a black hole: from an astrophysical point of view, black holes are in fact the next step after the cooling down below the "Schwarzschild's threshold" of a neutrons star (in this case, a very big one!). In the subsection 6.1.3 we discuss then how the apparent variation of the red-shift parameters as due to a time variation of this scale shows out as an accelerated expansion of the Universe.

In section 6.2 we consider the elementary matter excitations, which correspond to leptons and quarks, and the running of couplings. Particles exist as free states only in a perturbative limit. Furthermore, this limit is not the same for all of them. In order to explain the mass differences among particles, we would need a knowledge of the minimal entropy configuration more refined than an orbifold approximation: they are in fact tuned precisely by the moduli frozen at the orbifold point. In order to follow these details, we introduce and discuss the above mentioned "thermodynamical", statistical approach to the evaluation of the mass of elementary particles, and their couplings (smoothing down the target space to a differentiable manifold has the disadvantage of "projecting" onto a more classical configuration, and is therefore inappropriate for this purpose. In subsections 6.2.10 and 6.2.11 we comment on the "linearized representations" of the string vacuum, briefly discussing under what conditions, and up to what extent, the usual field theory running of couplings and masses, and the approaches based on the computation of the geometric probability of the phase spaces of particles, make sense).

In section 7 we come then to an explicit evaluation of masses, both of particles and of the bosons of the weak interactions, and the effective low-energy interaction terms. We discuss the degree of approximation under which these values are obtained, and give a rough estimate of the corrections they would receive if the string vacuum was known with a better accuracy. In particular, in section 7.5 we briefly discuss also baryon and meson masses. The investigation of the mass sector of the theory is completed in section 8, where we consider the mixing angles of weak decays (the Cabibbo-Kobayashi-Maskawa matrix) and CP violations. In our scenario neutrinos are massive; therefore, generation mixings and off-diagonal decays are expected to occur also among leptons.

Masses and couplings, as well as all flavour mixing and parity violation parameters, turn out to be given as functions of the age of the Universe. The quantity useful for a comparison with the values experimentally measured in accelerators or in general in a laboratory is therefore their present-day value. But knowing their behaviour along the history of the Universe allows us to test the predictions of this theoretical framework also in the case of astrophysical and cosmological observations.

In section 9 we consider the "Cosmic Microwave Background" radiation, and discuss how the existence of a $\sim 3^{\,0}$ Kelvin radiation comes out as a prediction in this framework. We discuss then also, in subsection 9.2, the case of dark matter. In our scenario, this is expected to not exist. We comment several cases which are usually considered to provide evidence for its existence, and propose how, within our framework, in each of them the effects attributed to dark matter receive an alternative explanation.

In section 10 we rediscuss then in the light of this proposal the constraints on the evolution of masses and couplings, coming from observations on ancient regions of the Universe, or, as is the case of the Oklo bound, from the history of our planet. We find out that the predicted behaviour is compatible with all the constraints. Not only, but in the case of the so-called "time dependence of α", it turns out to correctly predict the magnitude of the observed effect (section 10.1).

We conclude in 11, where we also comment on how general is the general validity of the functional 2.31, which seems to extend beyond critical superstring, selecting the actual configuration of the Universe over a set in which even configurations with generic speed of light are included.

2 The properties of a space-time *built* by light rays

2.1 The geometry of the Universe

We are used to consider the Universe as the set of things and phenomena that take place in a region of space-time we can observe, and therefore know of, thanks to the propagation of light rays. Intended as such, the Universe is an entity which possesses "intrinsic" properties, to which we can have a, somehow limited, sometimes partially distorted, access through the information carried to us by light [4]. In particular, the space-time is viewed as an "objective" frame, a geometric structure at least in principle independent of the way we get to know about it. For instance, the horizon of our observations is viewed as a "real" surface located at a distance corresponding to the age of the Universe. Here we want to discuss a different point of view, namely we are going to consider the space within the horizon as *built* by propagating light rays. This means that:

1. *all* the points of the Universe are causally connected to the observer. This means, not simply they fall *within* a space-like region, but *are* at a light-like distance, in space and time, from the observer. For the same reason,

2. these points are also light-connected to the origin, the "Big Bang" point.

As we are going to discuss, these assumptions, fully compatible with what we know about the Universe, lead to very dramatic consequences on the geometry of space-time, and the physics of the Universe, here practically identified with the way we perceive it as observers.

Let's consider the space-time corresponding to the region causally connected to us. This space is bounded by a horizon corresponding to the spheric surface, centered on our point of observation, whose radius is given by the maximal length stretched by light since the time of the Big Bang. Our attitude, the assumption at the base of our entire analysis, is that this region defines our "Universe": there is no space-time outside this region. This implies that the entire space-time originates from a "point".

At first look, the space included within the horizon looks more like a ball than a curved surface. If we set the origin of our system of coordinates at the point we are sitting and making observations, the Universe up to the horizon is by definition the set of the points satisfying the equation:

$$x_1^2 + x_2^2 + x_3^2 \leq \mathcal{T}^2 \,. \tag{2.1}$$

Nevertheless, there is today evidence that this space is curved, not only because of the presence of matter inside it, which obviously is a source for gravitation, and therefore for curvature: the evidence is that this space possesses a ground curvature. It is food for discussion whether this curvature originates from a kind of matter which escapes our detection, or anyway from processes that can be described in field theory. Here we want to discuss how, when restricted to our causal region, space-time has indeed a curved geometry, which exists,

[4] At least from a theoretical point of view, information is carried to us also from other "light-similar" rays, the gravitational waves. For historical experimental reasons however they are not as important as light rays.

so to speak, "before", i.e. regardless of, the physical processes that can take place inside it. More precisely, we will find that this condition on space-time is so strong that it will turn out to be not only at the origin of the curvature, but also of the existence of matter itself: in some sense, it is not matter that generates a curvature, but the curvature that generates matter. Let's go step by step, and see where does this curvature of space-time come from. Since we are going to consider an "empty" Universe, and therefore an initially flat space-time, $\mathcal{T}$ of equation 2.1 can be identified with age of the Universe itself (the speed of light in this empty space is c, and we use units for which $c = 1$). Owing to the finiteness of the speed of light, the region close to the horizon corresponds to the early Universe, and the horizon (i.e. for us the set of points $x_1^2 + x_2^2 + x_3^2 = \mathcal{T}^2$) effectively corresponds to the origin of space-time. This means that the points lying close to the horizon are indeed also close in space. The set of points $\vec{x} = (x_1, x_2, x_3) : \sqrt{x_1^2 + x_2^2 + x_3^2} \in [\mathcal{T}, \mathcal{T} - \epsilon]$ is, from an "objective" point of view, a ball of radius ϵ centered at the origin. We will see later that the minimal observable radius, the radius of what we call a "point" is in this scenario the Planck length. However, let's here for simplicity skip for a moment the question about whether really the origin is at $t = 0$ or, more appropriately, at $t = 1$ in Planck units, and therefore also whether at the origin the Universe is really a "point" or something with small but finite extension (for a horizon very large as compared to the Planck length, this approximation is justified). Let's here set the origin of the Universe, i.e. of space-time, at $(x_0 = t, x_1, x_2, x_3) = (0, 0, 0, 0)$. It is then clear that, from a "correct" geometric point of view, we, namely the observers, are sitting at a point on the hypersurface $x_1^2 + x_2^2 + x_3^2 = \mathcal{T}^2$. This defines a 2-sphere of radius $\mathcal{T}$.

The Ricci curvature scalar for a 2-sphere is given by $\mathcal{R} = \frac{2}{r^2}$, where r is the radius of the sphere, when thought as embedded in three dimensions. In our case, $r = \mathcal{T}$. In Ref. [2, 4] we derived the behaviour of the cosmological constant as a function of the Age of the Universe in an approximated way, $\Lambda \sim 1/\mathcal{T}^2$. It was not yet clear to us the non-perturbative framework in which to perform the exact computation. This will be explained in section 5.1, where we show how the normalization coefficient is 2: $\Lambda = 2/\mathcal{T}^2$. Let's consider the Einstein's equations:

$$\mathcal{R}_{\mu\nu} - \frac{1}{2} g_{\mu\nu} \mathcal{R} = 8\pi G_N T_{\mu\nu} + \Lambda g_{\mu\nu} \,, \tag{2.2}$$

Let's also consider the space as "empty", and therefore neglect the contribution of the stress-energy tensor. If we insert the value $\Lambda = 2/\mathcal{T}^2$, we see that the magnitude of the contribution of the cosmological constant to the curvature of space-time is exactly the one we expect, although it seems to be of the wrong sign. If we contract indices in eq. 2.2 with the inverse of the metric tensor, we obtain in fact:

$$-\mathcal{R} = 2\Lambda \,, \tag{2.3}$$

i.e.:

$$-\frac{2}{r^2} = \frac{2}{\mathcal{T}^2} \,. \tag{2.4}$$

The curvature is negative; this is however absolutely correct: the two-sphere we are considering is a surface oriented not outwards, as usual, but inwards; the observers sitting on this surface don't look outside but toward the center of the sphere, toward the "big-bang" point.

Therefore, the Ricci curvature *is negative*, as it must be: at the point the observer is sitting, this surface has a metric with hyperbolic signature. The three dimensional space is then built out of a series of shells, each one with the geometry of a two-sphere. Neighbouring points will in general sit on different shells, and will feel a different curvature: our hand is at a different distance from the horizon/big-bang than our eye, and therefore feels a different radius/age of the Universe. It should be clear that we are here giving up with space-time translational invariance. In this framework space-time is “absolute”: different points in space-time “see” in general a different horizon, and therefore also their measurements do not coincide. By knowing their location in the space-time it is however possible to relate the measurements of different observers. That’s all we can require: a “covariance” of the results.

We stress that, in order for this argument to make sense, we have to consider the space-time as something in expansion but indeed extended just up to the horizon. Otherwise, as is the case in the usual approach, we could not say that points which are close to the horizon, i.e. to the origin of time, are also close in space. In our case, at any time $\mathcal{T}$, there is no space beyond the horizon $x_1^2+x_2^2+x_3^2=\mathcal{T}^2$. Therefore, we cannot say that *today* we can *see* the past of something located at a point that was previously falling beyond the horizon. Not only is a journey of our viewing in space also a journey in time (the distant events we observe are past in time) but also the other way around is true: a journey in time is a journey in space. This means that a point that looks to be *there* (i.e. at a certain point in space), and of which we presume to see today the past, is in reality *not there*, the light comes indeed from somewhere else in space. The situation is illustrated in figure 1. In order to help the reader to visualize the situation, we show in figure 2 a two-dimensional, intuitive picture of the Universe, illustrating the fact that, both for a flat and a curved space-time, incident light rays arrive parallel to the hyperplane tangent to the observer, so that no difference is in practice locally observable (the only indication that the path of light is not straight but curved comes from a measurement of the cosmological constant or of a non-vanishing contribution to the stress-energy tensor). Owing to the curvature of space-time, the rays don’t come from the apparent horizon, the one obtained by straightly continuing the light paths along the tangent plane, but from a Planck-size horizon. Although useful to the purpose of illustrating how things are going, both figures 1 and 2 are slightly misleading, none of them being able to account for the real situation. In particular, from figure 1 we understand that there is a symmetry between picture A on the left and picture B on the right. The mapping from A to B, i.e. exchanging the origin with the horizon, involves a “time inversion”. This operation corresponds to a <u>duality</u> of the system. The configuration “B”, associated to the solution 2.4 of the Einstein’s equations, corresponds to one of the possible points of view, “pictures”, from which to look at the problem. Had we looked from the seemingly rather unnatural picture A, we would have concluded that the curvature is positive. However, this too is a legitimate point of view.

Let’s therefore have a closer look at the situation we are describing. As seen from the point of view of the origin of the Universe, the “surface” given by the equation $x_1^2+x_2^2+x_3^2=\mathcal{T}^2$ consists of points lying on non causally-related regions. We are sitting on a point of this surface, which however is an “ideal” surface: the only point we know to exist is the one at which we are sitting, the hypothesis of boundness of space-time being precisely justified

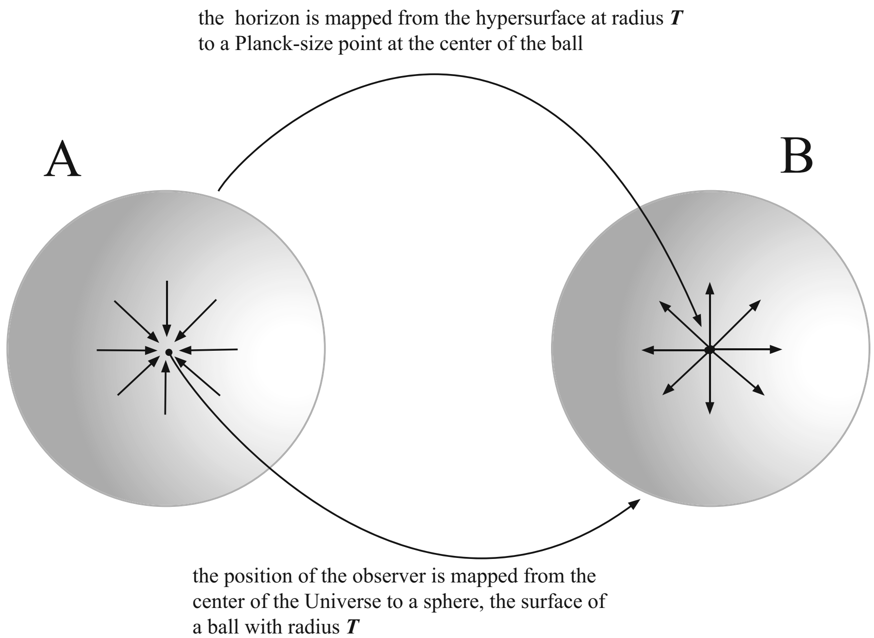


Figure 1: The ball A represents the Universe as it appears to us, located at the center and observing a space-time extended in any direction (solid angle 4π) up to the horizon at distance $\mathcal{T}$. The arrows show the direction of light from the horizon to us. Ball B on the right represents instead the "dual" situation, with the horizon, corresponding to the origin of space-time, at the center, and the light rays, indicated by the arrows, propagating this time outwards. We are sitting on a point on the surface, a two-sphere oriented inwards. The curvature therefore is negative. This surface has to be thought as an "ideal" surface: different points on the surface belong to different causal regions: there is no communication among them. The only point of it we know to really exist is the one at which we are located. From a "real" point of view, the two-sphere boundary of ball B is therefore a "class of surfaces", in each of which all the points of the surface have to be though to correspond to a single point in the "real" space-time. The path of light rays is therefore not straight but curved. Figure 2 helps to figure out the situation. Both figures should be taken in any case only as a hint, none of them being able to depict the exact situation.

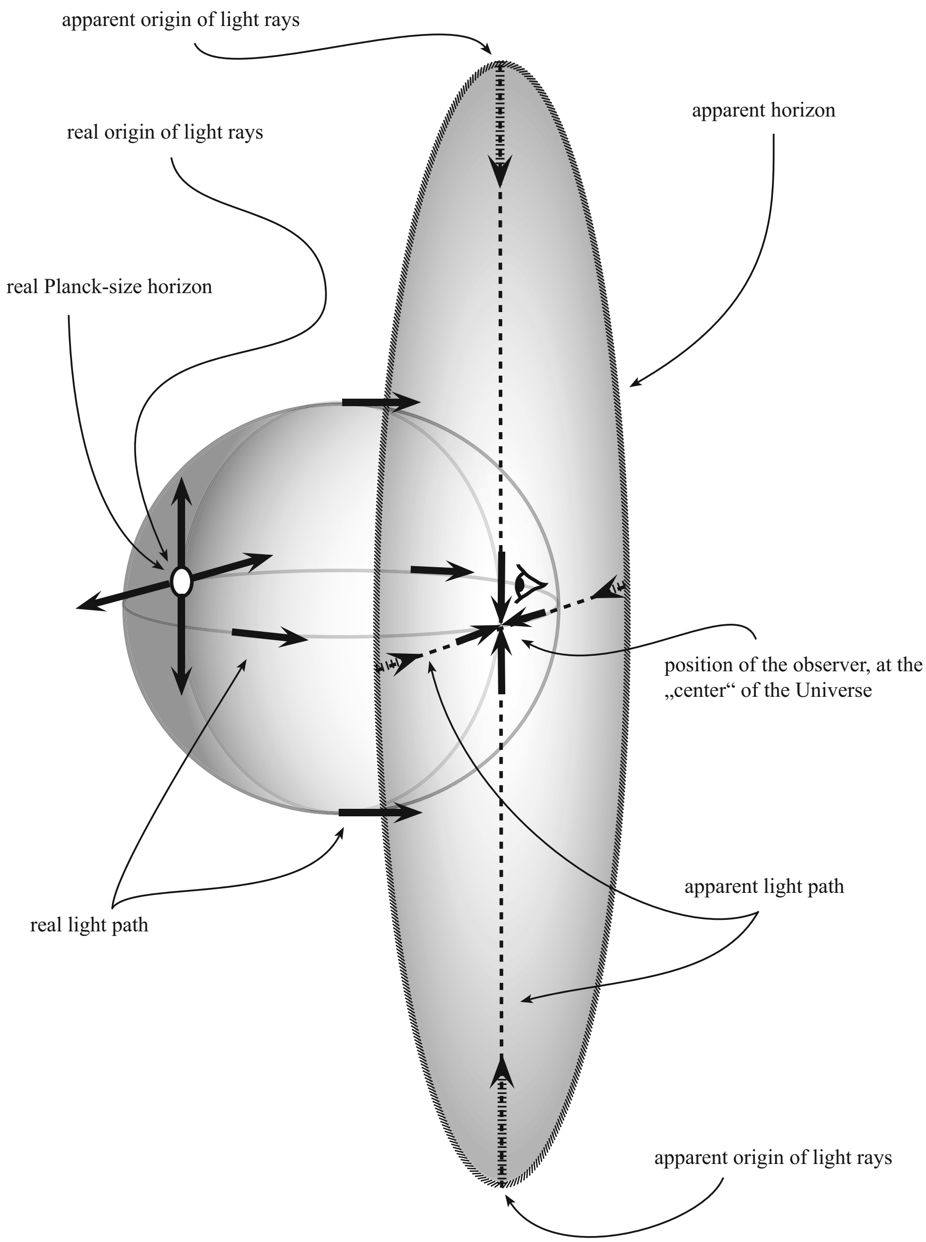


Figure 2: The disc represents the space-time as it appears to the observer: a flat, tangent space. The real space-time is however curved: the border of the disc corresponds to a Planck-size “point” on a 3-sphere, and is located diametrically to the position of the observer.

by the requirement of describing only the region causally connected to us (alternatively, we can think that this surface must be thought as equivalent to a "point"). In general, our Universe is the set of points, each one lying on a sphere $x_1^2 + x_2^2 + x_3^2 = r^2 \ < \mathcal{T}^2$, causally connected to us. Notice that any 2-sphere with radius $r < \mathcal{T}$ contains several points causally connected to us. The curvature of this set experienced by the observer is positive. In order to understand this, consider that this space corresponds to "shrinking" the various shells by different amounts: while the most external shell, the one on which the point of the observer is located, must be shrunk to a point, the neighbouring internal shells must be shrunk to progressively larger two-spheres. In other words, the space "opens up", as roughly illustrated in figure 3. It is not hard to realize that what we are describing is indeed the geometry of a 3-sphere. The curvature of a 3-sphere is three times larger than the one of a two-sphere

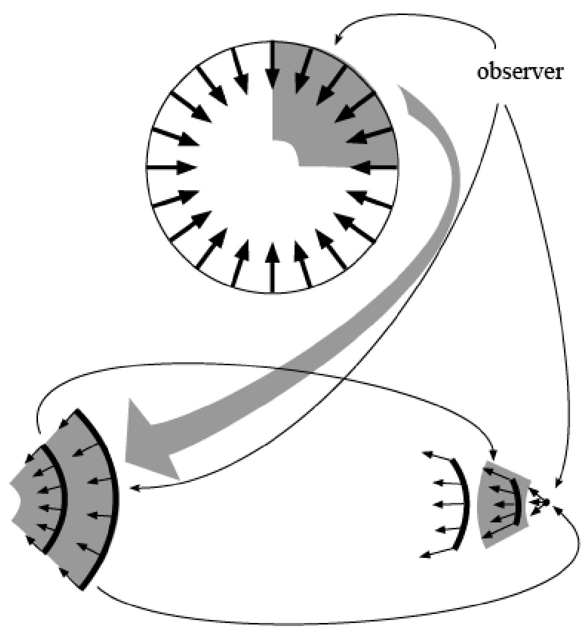


Figure 3: Shrinking to a point the two-sphere on which the observer is located (here represented by the external boundary of the disk), and considering the region causally connected to it, leads to a 3-sphere geometry and to a change of sign of the curvature.

with the same radius. This seems to imply that the cosmological constant only accounts for one third of the curvature of space-time. Where does it come from the missing part? This short discussion already shows that the problem is more complicated, and the solution deeper than just what we could expect from classical geometry. Indeed, the fact that we just encountered a first example of duality is a hint that the solution of the problem comes from thorough consideration of the consequences of the idea of considering our causal region as the total existing space-time, in association with the results of General Relativity. As we will see, these will be that space-time has a quantum nature. Let's for the moment skip the

details about getting the full contribution to the curvature. At this stage, we can already draw some qualitative conclusions. The fact that we measure a non-vanishing cosmological constant, and therefore that the curvature of space-time is non-vanishing, has the following origin. If we consider the region causally connected to us as the only one we indeed know to exist (and therefore have the right to consider), and therefore as the full existing space-time, it happens that the light rays starting from a "point", the origin of the Universe, end up also to a point, our point of observation. This is due to the fact that light has a finite speed, and therefore the region we see close to the horizon corresponds to the early Universe, located around the "big-bang point".

2.2 The Uncertainty Principle

A space-time as above described, i.e. limited by the natural horizon expanding at the speed of light, possesses, as we have seen, a non-vanishing curvature. The latter is due to the identification of the horizon with the origin of space-time. A non-vanishing curvature of space-time can be translated in terms of energy/mass: according to the Einstein's equations, it is in fact equivalent to the existence of a non-vanishing energy (mass) gap in the Universe, which acts as a source for the curvature. The Universe possesses therefore an energy, whose amount is related to the extension of space-time, or, equivalently to the age of the Universe itself. If we think at the Universe as a "bubble" of something blowing up out of "nothing" for a certain length in time and space, all that sounds much like the statement of a time/energy uncertainty relation: a fluctuation in space-time implies a fluctuation in energy, and suggests that indeed we can put these considerations on a more formal ground.

We will discuss here how the Heisenberg's Uncertainty Principle can be seen to precisely originate from this aspect of space-time, namely its built-in curvature produced by the finite speed of light. Put in other words, this means that what we observe to be the quantum nature of physical phenomena arises from the fact that all what we measure and observe, comes to us mediated by light (or gravity, that from this respect behaves (i.e. propagates) similarly to light). To simplify, we could ultimately say that the world shows a "wave-like behaviour" basically because we experience it through a wave-like medium [5].

Historically, the Uncertainty Principle was stated for the point-like particle, not for the Universe as a whole: the basics of Quantum Mechanics have been first established for micro-phenomena, not for macroscopic ones, although, as we will discuss, they manifest themselves also at a cosmological scale (for instance, through the existence of the so-called cosmological constant [6]). To make contact with the usual formulation of the Uncertainty inequalities, we must first clarify what in this framework a point-like particle is. We anticipate here that, at the end of the analysis, we will end up with the existence of a minimal length in the Universe, the Planck length, that must therefore be considered also the "size" of a point. However, here we don't want to make any hypothesis about the existence of a minimal length: we will

[5]See Ref. [1] for a discussion of the subtleties related to the probabilistic interpretation of dynamics usually associated to Quantum Mechanics.

[6]Considerations somehow similar to those we expressed in Ref. [2, 4], relating cosmological constant and Heisenberg's Principle, are to be found also in Ref. [5].

just say that a point particle is an object with a certain "radius", R_0, that can be zero or of finite size. Of course, also the "classical" approach is included in the discussion, because it corresponds to the case in which this radius is zero. Since it turns out to be convenient to work in Planck units, to start with we make the hypothesis that this radius is precisely 1 when measured in Planck units. This will turn out to be the "right" solution, but for the moment it is just a convenient value to start with the discussion, that we will generalize to any smaller size.

Let's consider a "particle" $\mathcal{P}$ which exists for a time $\mathcal{T}$ and then decays. From the point of view of this particle (i.e. in its rest frame), during its existence there will be an "effective Universe" which opens up for it, with horizon at distance $\mathcal{T}$. We can consider this particle to be the "observer". For the reason explained above, this universe will be curved. We can consider that the curvature comes entirely from the gravitational field induced by the rest energy of the particle. In this case, the Einstein's equations tell us that the curvature of space-time is precisely the stress-energy contribution to the curvature at a certain point on the "surface" of the particle, if we imagine it as a ball of Planck-size radius:

$$\mathcal{R}_{\mu\nu}(R) - \frac{1}{2} g_{\mu\nu} \mathcal{R}(R) = 8\pi G_N T_{\mu\nu}(R, m_{\mathcal{P}}) \,, \tag{2.5}$$

where $R > R_0$, and we have absorbed the Λ-term into a redefinition of the stress energy tensor. If we can go to the rest frame of the particle, then the above equation simplifies because the only non-vanishing component of the stress-energy tensor is T_{00}. However, we encounter a problem: what is the meaning of "rest" frame, if we are going to discuss about an uncertainty in energy and time, and therefore, also in momentum/position? A fluctuation in time is also a fluctuation in space, and indeed, the uncertainty relation implied by Eq. 2.5 is an overall uncertainty, that accounts for all uncertainties contributing at the same time. A "boost" to the rest frame corresponds therefore to an expansion of space coordinates, so that, at the end, the particle feels the curvature just along one coordinate (whether space or time-like it is irrelevant, because the space and time intervals we are considering correspond to the extension of horizon of space-time. The latter is a light-like surface, in which the spatial radius is as large as the extension in time).

In the following, we will proceed by considering a Universe expanding along 3 + 1 dimensions, as actually is. However, the existence of a minimal energy, in a bounded Universe as described, is not constrained to four space-time dimensions: we would get the same conclusions starting from any space-time dimensionality, by considering the embedding of the Einstein's equations into higher dimensions. The Uncertainty relation, and the conclusion about the quantum nature of space-time, is a general property. Indeed, as we partly saw in Ref. [2] and will rediscuss along this paper, under the conditions that arise in the scenario we are presenting, four space-time dimensions are automatically selected among all the possible space-time configurations of String Theory [7], a theory naturally defined in a higher number of dimensions. For the rest of this paragraph we will therefore present our discussion for the case of four space-time dimensions, but it can be easily generalized to higher dimensions.

[7] We will see that this "selection" does not occur in the classical sense of solution of an equation, but in the quantum sense of "statistically favoured".

By contracting indices in the "rest frame", from 2.5 we obtain [8]:

$$-\mathcal{R} \;=\; 8\pi G_N T_{00} \;=\; \frac{1}{\mathcal{T}^2}\,, \tag{2.6}$$

We will not mind here about the sign of the curvature: this depends on the orientation of space-time, that, as seen "from the particle", has opposite orientation than as seen "from the observer located at the horizon" (with reference to figure 1, it is a matter of passing from picture A on the left to picture B on the right. Although locally the absolute value of the curvature remains the same, the orientation of space gets inverted, and one passes from the local geometry of a sphere to the one of a hyperboloid).

Obtaining the term T_{00} is not so trivial: the introduction of a minimal length implies in fact a discretization of space. A discretization cannot be obtained through balls, that would leave "empty spaces", but through cubes, "cells". Therefore, although it is convenient to imagine the particle as a ball, this picture probably does not really correspond to what a "point" in this space is. It is perhaps more appropriate to think in terms of "elementary cells". To get the term T_{00} we can start with a ball $\mathcal{B}$ made up of a huge number of cells, and then take the limit for the number of cells going to one. In this case, we can use much of what we know about the behaviour of the gravitational force. The total energy, i.e. the integral $\int d^3x T_{00}(\vec{x})$ over the volume of the ball, is given by the quantity $\int d^3x_1 d^3x_2\, G_N \rho_1\rho_2/|\vec{x}_1 - \vec{x}_2|$, where also the integration on dx_2 is performed over the volume of the ball. The result is:

$$\int_{\mathcal{B}} d^3x\, T_{00}(\vec{x}) \;=\; \frac{1}{4\pi} G_N \frac{\mathrm{M}^2}{R}\,, \tag{2.7}$$

where ρ is the mass density of the particle, M its mass, and R the radius of the ball. The term T_{00} is then:

$$T_{00} \;\sim\; \frac{1}{4\pi} G_N \frac{\rho \mathrm{M}}{R}\,. \tag{2.8}$$

For an extended homogeneous ball, the density is the mass divided by the volume $V = \frac{4}{3}\pi R^3$. The case of a true point particle is obtained by taking the limit of zero radius: the density ρ becomes then the mass itself multiplied by a delta function centered at the point where the particle is located. In our case, points of space-time are promoted to Planck-size cells, so that the ball has to be thought as an object made out of many, say "n", cells $\mathcal{C}$. The delta function of above becomes a kind of step function ϑ supported on the cell $\mathcal{C}$:

$$\rho(\vec{x}) \;=\; \mathrm{M}\,\delta(\vec{x}) \quad \to \; \mathrm{M}\,\vartheta(\mathcal{C})\,. \tag{2.9}$$

If ρ is the mass of one cell, the unit of volume, the total mass is $\mathrm{M} = n\,\rho$. The integral 2.7 becomes in this case:

$$\int_{\mathcal{B}} d^3x T_{00}(\vec{x}) \;\to\; \frac{1}{4\pi} G_N \frac{(n\rho)(n\rho)}{R}\,, \tag{2.10}$$

[8] g^{00} can be set to 1 by an overall rescaling of the metric, that basically amounts to a choice of units for the speed of light.

and the limit of "zero radius" becomes now the case $n = 1$. The contribution of the Planck-size "point particle" is therefore:

$$T_{00} \; = \; \frac{1}{4\pi} G_N \frac{M^2}{(R=1)} \, . \tag{2.11}$$

Inserting this value in eq. 2.5, we obtain $\sqrt{2}\,\mathrm{M} \; = \; 1/\mathcal{T}$ $(G_N = 1)$. This value tells us that, during a time interval $\Delta t \; = \; \mathcal{T}$, the system undergoes an energy fluctuation $\Delta E \; = \; M$, such that $\Delta E \Delta t \; = \; \mathrm{M} \times \mathcal{T} \; = \; 1/\sqrt{2}$ (in units $\hbar \; = \; 1$). This is the minimal energy fluctuation produced by the existence of the particle, and therefore saturates the bound as an equality. The normalization is however not quite the one of Heisenberg, that at the saturation reads: $\Delta E \times \Delta t \; = \; 1/2$. The mismatch by a factor $1/\sqrt{2}$, basically amounting to a redefinition of the Planck constant $\hbar$, is due to a rather subtle property of the geometry of space-time. We will see later that indeed space-time is an "orbifold", and the experimental value of the fundamental constants is measured in the actual space-time, that feels the coordinate contraction due to this projection. Once taken into account, the correct normalization pops out precisely a factor $1/\sqrt{2}$. The generic Heisenberg's inequality is then a consequence of the fact that what we have considered till now is just the "ground" contribution, the one given by the bare mass of the particle, not considering any other kind of energy contribution.

An important ingredient in the above derivation of the Uncertainty relation is the existence of a minimal observable length, the Planck length. We already anticipated however that this assumption is not necessary. Let's see what happens if we relax the condition that the minimal length must be identified with the Planck length. For a generic "radius" R, i.e. for a different size of the unit cell, the density gets rescaled as the cube of the ratio of the two units:

$$\rho \; \to \; \rho' \; = \; \rho \, \times \, \left(\frac{(R=1)}{R'} \right)^3 \, . \tag{2.12}$$

For a generic "radius", the stress-energy contribution of above is $T_{00} \; = \; \frac{1}{4\pi} G_N \mathrm{M}^2 / R^4$. Therefore, according to our derivation, the Heisenberg's inequality should in general read:

$$\Delta E \, \Delta t \; \geq \; \frac{1}{\sqrt{2}} \frac{R_0^2}{G_N} \, , \tag{2.13}$$

where R_0 is the minimal observable radius, i.e. the radius of what we call a point-like object (not to be confused with the uncertainty in its position! this radius is universal and independent on the uncertainty in the momentum). Let's now suppose we can observe intervals shorter than 1 in Planck units: consider the possibility of observing a particle (in its rest frame) for a time Δt such that $R_0 \; < \; \Delta t \; < \; 1$ in units $G_N = c = 1$. According to 2.13, the energy fluctuation during this time is then:

$$\Delta E \; > \; \frac{R_0}{(G_N = 1)} \; = \; \frac{1}{\sqrt{2}} R_0 \, . \tag{2.14}$$

As we already pointed out, owing to the orbifold nature of space-time, the space coordinates will prove to be renormalized by a $1/\sqrt{2}$ scaling factor. Once the correct coordinate normalization is taken into account, we have a factor $1/2$ instead of $1/\sqrt{2}$. Such a rest-frame energy fluctuation is confined within a region of radius R_0 (that we can therefore consider as an upper bound for the Schwarzshild radius), and it appears as a "mass fluctuation". This "mass" is however larger than $R_0/2$, and therefore is also larger than the Schwarzschild radius of the particle: this particle is therefore a black hole, and cannot be observed. In this framework, the existence of a minimal observable length and its identification with the Planck length turn out therefore to be consequences of General Relativity, not independent assumptions, and the Uncertainty Principle in its usual formulation the only possible inequality.

Although we have presented our arguments for a four dimensional space-time, it is easy to recognize that they can be straightforwardly generalized to a higher number of space-time dimensions. What basically changes is a normalization, coming from contractions and curvature terms. However, all this can be reabsorbed in a redefinition of the Newton constant in higher dimensions (or, equivalently, of the Planck mass). By consistency, this must reduce, upon compactification, to the expressions we just derived: any higher dimensional extension of General Relativity must in fact reduce to the usual one upon compactification to four dimensions.

To summarize:

An Uncertainty relation is the consequence of General Relativity and of the properties of propagation of light, namely its finite speed. All what we know about the Universe comes to us through light rays (or gravitational fields, the only two long range interactions, both propagating at the speed of light). Boundness of the Universe alone however does not imply the existence of a minimal energy, i.e. a maximal measured wavelength. Essential for this is the boundary condition of space-time, and precisely the fact that the surface at the horizon corresponds indeed to a "point", the origin, intended as explained above. This "closes" the space, implying that points at the boundary surface are "identified", and produces a curvature. In this way, a segment becomes a circle, a flat space a sphere. The existence of a minimal length, identified with the Planck length, implies on the other hand a change of perspective in the approach to the geometry of space-time: in this perspective, differential geometry turns out to be only an approximation, that works well only at "large" scales: at the small scale, there exist no points intended as objects with no extension. By looking back at figure 1, we can now see that in passing from picture A to picture B, there is no singularity in mapping a "point" into a two-sphere, and a two-sphere into a disk, as in figure 2, because a point is not a "point", and a two-sphere without the "point" is indeed a disk. Therefore, in this scenario the identification of the boundary surface with the "point at the origin" is an operation that makes sense.

We have seen that the Uncertainty Principle is not an external input, being already "built in" in a space-time bounded by and enclosed within the horizon set by the propagation of light. In such a space-time, everything what happens appears to possess a "wave-like" behaviour, because it comes to the observer through light waves. In particular, the boundary conditions of this space (or, equivalently, the geometry of light propagation) imply the existence of a non-vanishing curvature for any finite extension in space and time; this implies in turn the existence of a minimal energy/momentum, related to the time/space extension. These conditions say that the Universe, intended as the "bare" space-time and all what is inside it, possesses a "quantum" nature. Indeed, the inequality, or better the set of inequalities (energy/time and the relativistically related momentum/position inequality) 2.13 state the wave-like behaviour of matter. The theoretical implementation of this behaviour requires the substitution of the Poisson brackets with commutators, leading to quantum mechanics. Quantization of space and time means that also gravity is quantized.

We have seen that below the Planck length the Universe is no more observable in the ordinary sense because it behaves like a black hole. On the other hand, if we invert the coordinates with respect to the Planck length, we pass from the outside to the inside of the black hole. There, density is no more critical, and we see objects and "particles", that were hidden to us by the Schwarzschild horizon. What previously, i.e. from outside, were objects and particles, appear now as black holes. A theory that quantizes the Universe must therefore possess a "T-duality", which, by inverting coordinates in Planck length units, in practice exchanges inside with outside of the "Planck-size surface". Owing to this symmetry, the Planck length becomes automatically the minimal effective length. The mapping between A and B of figure 1 is indeed a T-duality. Under inversion of the space-length with respect to the Planck scale, it maps the "over-Planckian" region, i.e. from radius 1 to radius $\mathcal{T}$, to the "sub-Planckian" region, from radius $1/\mathcal{T}$ to 1. In this way, it "unfolds" the Universe as a black hole, and we pass from looking at it from inside to "outside".

A theory with these requirements exists: it is String Theory [9], which possesses a built-in T-duality relating the observed world with a dual world of black holes (the massive string states with mass above the string/Planck scale [10]). As is known, String Theory does't possess a built-in mechanism allowing to select one, or a family, out of the huge amount of its possible configurations, each one in principle describing a "universe" with its own time evolution. We will find that the whole phase space of configurations indeed does possess such a selection mechanism, based on the fact that not all the configurations have the same occupation in the phase space. Some are more frequent than other ones, and the resulting

[9] From now on, we will consider String Theory as "the" candidate satisfying these requirements. This choice is not the consequence of a not (yet?) existing uniqueness theorem, stating that the approach of string theory is the only way of quantizing gravity that possesses these characteristics. To our purpose, it is sufficient that String Theory constitutes a good approximation of the physics we want to describe.

[10] A priori, the string scale does not necessarily correspond to the Planck scale: it does only under particular conditions. As we will discuss in the next sections, these conditions will turn out to be satisfied by the string theory solution eventually selected by the dynamics of the Universe under the conditions we are considering. For simplicity, we therefore consider right now the two scales to be equivalent.

universe will appear more like the first ones than like the second ones. As we will discuss, in the favoured configuration a four-dimensional space time, as well as the breaking of T-duality (and supersymmetry), are automatically implied, something a priori not obvious. In order to work it out, we must preface some considerations.

As we discussed in Ref. [2], quantization of space and time involves a deep change of perspective in the treatment of dynamical problems. What in fact we usually do, is to write, or at least to think in terms of, equations that should give the evolution of the system along a special coordinate, the time. Here however time is a field, and we cannot look for a Lagrangian formulation of a theory that must give us also the evolution of the time itself, intended as a field, a function of a time parameter. Here I have presented the problem in a "brute" way. Usually, this question does not arise in this way: the problem appears hidden under other forms: one looks for a correct Lagrangian formulation for a theory that contains as a degree of freedom not the time (and space) itself, but the graviton, a field that encodes the geometry of space and time. Once the problem is phrased in this way, it seems possible to look for a "traditional" Lagrangian formulation. However, I stress that this can be a misleading approach, in that it hides the true, fundamental problem, namely that of being time and space themselves quantized (a hint of this is the existence, discussed in the previous paragraph, of a minimal length, the Planck length).

In order to understand the "laws of dynamics" of this quantum system, we begin by observing that, at any volume V, the configuration that occurs the higher number of times, the one in which the system "rests the most of time" and therefore "weights" more, is the one corresponding to the minimum of entropy. In order to see this, let's start with an example on a very simple system. Let's consider the typical case of a gas of identical particles, confined in box separated in two sectors, A and B, by a removable wall. At the starting point, the gas is entirely contained in part B, as illustrated in figure 4, picture 1). If we remove the wall and let the gas to expand, the entropy of the system increases. Situation 1) is therefore less entropic than situation 2). For the same reason, also the configuration 3) is less entropic than 2). Indeed, 1) and 3) are equivalent. In the system of interest for us, the Universe, owing to an obvious symmetry configuration 1) and 3) are indeed indistinguishable.

Let's come now to our system, the Universe, which certainly is not simply a "gas of identical particles". Nevertheless, owing to the existence of a minimal length, and to the fact that for any finite volume V there is also a minimal energy/momentum step, it is possible to consider a "discretization" of the (phase) space into cells, and think that a configuration with a certain entropy occupies a certain number of cells. By increasing the volume of space, and therefore also of phase space, we increase the number of available cells, and with this also the number of possibilities for this configuration to be realized. If we keep fixed the distribution of particles and energies, we observe a decrease of entropy, as a consequence of the higher probability concentration. However, there is an increased number of equivalent configurations, in the sense of 1) and 3) of figure 4. The phase space of our system should somehow be thought as a sort of "lattice", as in figure 5, where the region of phase space occupied by the actual configuration is coloured in grey. Equivalent configurations correspond to a permutation of the position of the grey square. Owing to the multiplicative nature of phase space, it is not hard to realize that the "weight" of a

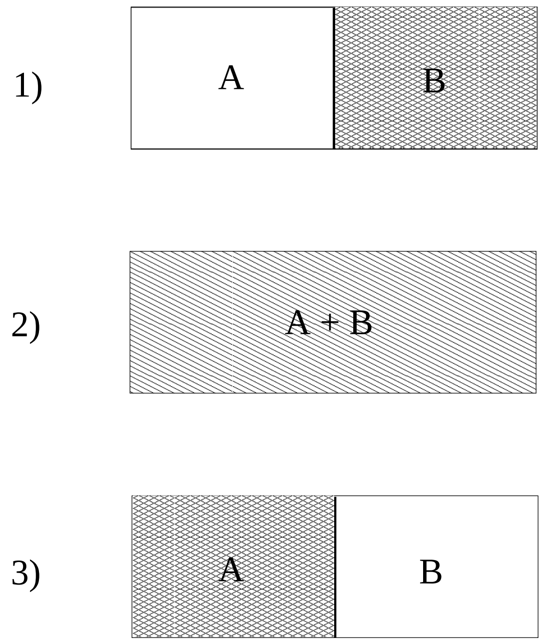


Figure 4: The adiabatic expansion of a gas of identical particles. owing to the symmetries of space-time, situations 1) and 3) are equivalent.

configuration ψ is then proportional to:

$$\mathrm{weight}(\psi) \;\propto\; \prod_{A_i} P_\psi(A_i)^{P_\psi(A_i)}\,, \tag{2.15}$$

where A_i are the states the system consists of. A way to see this is the following. We can view any configuration as being obtained from a "maximal one" through (a chain of) symmetry reductions. Let's consider the Universe at space-time volume V, and the corresponding phase space, with "symmetry group" $G_V = \prod_i G_{V_i}$ and volume $\mathcal{V}_i(G_{V_i})$. "Reduced" configurations consist of products of spaces corresponding to the states A_i, obtained by dividing G_V by some subgroup $Q = \prod_i Q_{v_i}$, such that we are left with a symmetry group $G' = G_V/Q$, $G' = \prod_i G'_i$, whose factors have volume v_i. The volume $\|Q_{v_i}\|$ of Q_{v_i} is $\mathcal{V}_i/v_i$, which is also the inverse of the probability of A_i in the reduced configuration ψ' as compared to the maximal one: $P(A_i) \propto v_i/\mathcal{V}_i = 1/\|Q_{v_i}\|$. Dividing by a symmetry group generates a larger phase space of configurations: after quotientation, the phase space consists of all the equivalent configurations corresponding to the orbits of the group Q, and its volume gets consequently expanded: $\mathcal{V}_i \to \mathcal{V}_i^{\|Q_{v_i}\|}$. The weight of each subspace of ψ' is that of a <u>product</u> of configurations with equal probability. Namely, the configuration is given by the product $\prod \psi'_{v_i}$, $\psi'_{v_i} \in \{\psi'_{v_i} = Q_{v_i}\psi_{v_i}\}$, and follows the rules of composite probabilities. In order to keep consistent the normalization of probabilities through the set of configurations generated by quotientation, we must therefore re-normalize, reducing back volumes to the

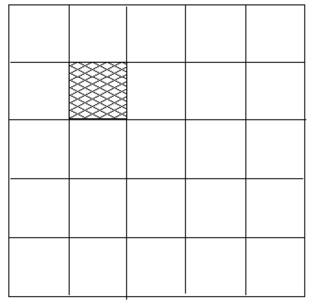


Figure 5: At any volume V the phase space of the Universe can be viewed as a discrete "lattice" with certain symmetries. The shadowed cell represents a configuration concentrated on a certain region of the phase space, symmetric to the other ones.

initial size, by taking the $\|Q_{v_i}\|$-th root of each subvolume. This is equivalent to raising each probability to the $1/\|Q_{v_i}\|$-th power, that is, to the power $1/(1/P(A_i)) = P(A_i)$. Expression 2.15 follows therefore, up to an overall normalization of probabilities, which does not affect our conclusions. These considerations are supported by the fact that the theory we are going to consider is String Theory. Under the hypothesis of uniqueness of string theory, any string configuration can be viewed as obtained through a process of (super)symmetry reductions, starting from a configuration with the highest super(symmetry). We will make an extensive use of this property in the next sections, where we will consider a class of orbifold constructions, as a base from which to start in order to approximate the string configurations. This class of constructions can be considered as corresponding to a subspace of the entire phase space, in which some parameters (moduli) are frozen, and the full symmetry group has been partially reduced from the very beginning, in order to make easier an explicit solution. However, the fact that string configurations are obtained via symmetry reduction remains true also out of the orbifold point. Simply, in the most general case we don't mod out with a discrete group.

Since $P(A) < 1$, it is always $P(A)^{P(A)} > P(A)$. According to these considerations, we introduce the "partition function" of the system, defined, for any finite volume V, as:

$$\mathcal{Z}_V \equiv \int \mathcal{D}\psi \prod_{A_i} P_\psi(A_i)^{P_\psi(A_i)} , \tag{2.16}$$

where the sum (the "integral") is performed over all the possible configurations of the system. Let's now introduced entropy, defined as usual by:

$$S_i \equiv - \ln < P_i > . \tag{2.17}$$

According to this definition, more entropic configurations are those in which the probability is more spread out, less entropic ones those in which the probability is more concentrated. According to this definition, 2.16 can also be expressed as:

$$\mathcal{Z}_V \equiv \int \mathcal{D}\psi \; \mathrm{e}^{-S(\psi)} , \tag{2.18}$$

This expression is quite reminiscent of the path integral. In this case, instead of the paths we have the configurations (that we can consider as "paths", "points along the history" of the Universe), weighted with the exponential of entropy. It is easy to recognize that the dominant configuration is the one of minimal entropy [11].

The question now is: how does the system evolve? A quantum system does not evolve according to differential equations, but according to the laws of the probability of its configurations. Let's consider the set of all $\mathcal{Z}_V$, obtained by letting the volume V assume any possible value. As we observed, for any volume V the dominant configuration is the one of minimal entropy. It is important here to remark that, when we speak of a configuration at volume V, we intend a configuration whose space-time support $\mathcal{C}$ has volume V. This means, it is not supported in $\mathcal{C}' \subset \mathcal{C}$, with $V(\mathcal{C}') < V(\mathcal{C})$. According to this definition, the configurations at volume V are not a subset of the configurations at larger volumes. If we indicate with $\psi_V^{\min}$ the configuration of minimal entropy at volume V, we have that:

$$S(\psi_{V+\delta V}^{\min}) \ > \ S(\psi_V^{\min}) \,, \tag{2.19}$$

i.e. entropy of the minimal configuration increases with volume. This is easy to understand, because with the volume increases also the support, and therefore the spread, of a configuration. As a consequence, the "weight" of the minimal configuration decreases as the volume increases. However, in comparison with the weight of the higher entropy configurations at the same volume, it increases. This means that, as volume increases, the configuration of minimal entropy becomes the more and more favoured over the other ones. This is a simple consequence of the fact that, with increasing volumes, the combinatorics of phase space too increase (this issue has some subtleties and will be discussed more extensively in section 4.1).

If we consider the "evolution" of the system through the set of volumes $\{V\}$, we observe a "continuity" of the configuration, a kind of "conservation", according to which, in passing from V to $V+\delta V$, $\psi_V^{\min}$ mostly "flows" to $\psi_{V+\delta V}^{\min}$, because this is the configuration at volume $V+\delta V$ with the highest projection probability on $\psi_V^{\min}$ (it is the closest one). More concretely, the configurations $\psi_V^{\min}$ and $\psi_{V+\delta V}^{\min}$ will be expressed as a superposition of the states at volume $V+\delta V$:

$$\begin{aligned} \psi_V^{\min} &= \sum_i \Psi_V^i |A_i> \,, \\ \psi_{V+\delta V}^{\min} &= \sum_j \Psi_{V+\delta V}^j |A_j> \,, \end{aligned} \tag{2.20}$$

[11] In Ref. [1] we present another approach, in which instead of considering each configuration as made out of its "micro-states", we consider probabilities of whole configurations in the full phase space, thereby obtaining dual formulae to 2.15 and 2.18, in which the entropy considered is not the sum of the microscopical entropies of the states of a configuration, but expresses the occupation of a configuration as an element in the full phase space, the one including all the orbits of the symmetry groups. We distinguish these two definitions of entropy as $S_{\rm micro}$, the one considered in this paper, and $S_{\rm macro}$. The weight of a configuration in the full phase space is proportional to $e^{S_{\rm macro}}$. At any fixed volume V, microscopical and macroscopical entropies turn out to be opposite to each other: $S_{\rm micro} = -S_{\rm macro}$, up to a constant shift, thereby ensuring the equivalence of the two approaches.

where it is however intended that, by definition, $\psi_V^{\rm min}$, although expressible in terms of the states at volume $V + \delta V$, does not belong to the set of configurations $\{\mathcal{C}(V+\delta V)\}$. The projection of $\psi_V^{\rm min}$ onto $\psi_{V+\delta V}^{\rm min}$ is then:

$$\frac{|<\psi_V^{\rm min}|\psi_{V+\delta V}^{\rm min}>|^2}{|<\psi_{V+\delta V}^{\rm min}|\psi_{V+\delta V}^{\rm min}>|^2} \; = \; \frac{|\sum_{ij}\delta_{ij}\Psi_V^{*\,i}\Psi^j_{V+\delta V}|^2}{|\sum_j \Psi^{*\,j}_{V+\delta V}\Psi^j_{V+\delta V}|^2} \, . \tag{2.21}$$

This is clearly higher than the projection on higher entropy configurations of $\{\mathcal{C}(V+\delta V)\}$, which are "less concentrated" on $\psi_V^{\rm min}$. Along the sequence of increasing space-time volumes, the system evolves therefore by minimizing at any step the increase of entropy.

In section 3 we will see that the minimal entropy configuration contains massless degrees of freedom (the photon and the graviton), that freely expand at the speed of light, and "stir" the horizon, increasing the volume V. In general, this would not yet necessarily imply an effective "expansion" of the Universe, intended as something really distinguished from a contraction, at least as long as T-duality remains unbroken. However, another feature of the minimal entropy configuration is also that T-duality is broken. This implies that the system evolves toward increasing space-time volumes, and increasing entropies. The absolute minimal volume/entropy configuration is realized at volume 1 in Planck units. This is the starting point of the evolution within the class of minimal entropy configurations. Since this is the class of dominant configurations, that turn out to dominate the more and more over the other configurations, we conclude that the Universe looks the more and more like the minimal entropy configuration, characterized, as we will see in section 5, by a pressureless, non-accelerated expansion of the Universe, in which time reversal and T-duality are broken. It is in this sense that it is possible to order the history of the Universe along the time coordinate. At large times (or equivalently space-time volumes) we are allowed to talk about this as "the configuration" implied by 2.18. Within the class of minimal configurations, it is then possible to identify an arrow of the evolution, and entropy can be put in bijection with time. This is the "quantum gravity" version of the second law of thermodynamics: in this framework it appears to be related to the breaking of parity and T-duality.

Under these circumstances, it is possible to establish a contact with the usual approach to quantum mechanics and thermodynamics. In a space-time of finite volume it is in fact only after the breaking of T-duality that it is possible to define an effective action analogous to the usual one, i.e. through a Lagrangian formulation (at finite space-time volume, an expression of the integral over space-time such as the ordinary one, taking into account only the light, sub-Planckian modes, would explicitly break T-duality along space-time coordinates). Since in the sequence of the most favoured configurations entropy never decreases, and a minimal entropy configuration preferably flows (in the quantum probabilistic sense) to the minimal entropy configuration at the subsequent "time", we can speak of a "time-dependent global solution" of minimal entropy, $\psi^{\rm min}$, consisting of the sequence of solutions $\psi^{\rm min}(\mathcal{T})$ of minimal entropy at any time: $\psi^{\rm min} = \left\{\psi^{\rm min}(\mathcal{T})\right\}$.

Strictly speaking, a string configuration with a certain entropy is a certain vacuum, characterized by a spectrum, i.e. a certain content of states and fields, which are functions of certain parameters, the "coordinates" of their position in space and time. In general, the string space is divided into an internal, compact space and an extended one, more

properly known as the space-time as such. The spectrum is determined by the "geometry" of the "internal" string space, while the geometry of the space-time is in turn modified by the matter and field content. In our scenario, the entire string space is compact, and the distinction between the two relies on whether there exist or not states propagating along a certain subspace of the target space or not. We will see that, in the dominant configuration, this happens only along four dimensions. Entropy depends not only on the geometry of the internal space. but also on the distribution of matter and fields in the space-time (this latter is the entropy in its usual formulation). Although the sum in 2.18 counts with a specific weight, and therefore distinguishes, also among almost identical configurations, differing for instance by a slight distinguished shape of a galaxy cluster, to the purpose of our analysis it is convenient to group the string configurations into "classes", corresponding to a certain spectrum, i.e. distinguished according to their internal configuration: as we will discuss in section 4.1, a difference in the distribution of matter and fields in the space-time leads to variations of entropy which are smaller than those produced by changes in the configuration of the internal space (such as un-twisting of coordinates etc...). The contribution to 2.18 within each such class is almost the same, but also the way the Universe appears is very similar. It is in this sense that we refer to $\psi^{\min}(\mathcal{T})$ as to a "class" of configurations. In the following, this will be always understood in this way, even when it is not explicitly specified.

⋆⋆⋆

Along the class of minimal entropy configurations, 2.18 reduces to the ordinary path integral [12]. In order to see this, consider the well known thermodynamic relations:

$$dS \,=\, \frac{dQ}{T}\,, \tag{2.22}$$

where T is the temperature of the system, and:

$$dQ \,=\, dU \,+\, PdV\,, \tag{2.23}$$

also known as second and first law of thermodynamics respectively. As we will see in section 5, along the sequence $\left\{\psi^{\min}(\mathcal{T})\right\}$ the universe in its whole can be viewed as a Schwarzschild black hole. Its temperature is then proportional to the inverse of its age, according to the relation:

$$T \,=\, \frac{\hbar c^3}{8\pi G M k} \tag{2.24}$$

[12] As discussed in Ref. [1], through quantization, i.e. basically through the implementation of the Heisenberg's Uncertainty Relations, one provides the "classical theory", corresponding to the dominant configuration of the universe, with a way of accounting for the contribution to the weighted sum of *all* the other configurations. Choosing to describe the physics of the universe through its dominant configuration, plus a quantization principle, can be viewed therefore as a way of separating the solution into a "mean value approximation" plus a perturbation term which accounts for the deviations from the central value.

where k is the Boltzmann constant and M the mass of the Universe, proportional to its age according to $2GM = \mathcal{T}$. As the universe is in our framework by definition "ab-solute", its evolution is a pressureless expansion [13]. In this class of configurations, the second term on the r.h.s. of Eq. 2.23 therefore vanishes, and we can identify the heat with the total energy. We obtain therefore the relation:

$$dS = \frac{dU}{T}, \tag{2.25}$$

between entropy and energy of the system at a certain temperature (or, equivalently, age, "point in the history of the Universe"). By substituting energy and temperature to the entropy in 2.18 according to the above relation, we get:

$$\mathcal{Z} \equiv \int \mathcal{D}\psi \, \mathrm{e}^{-\int \frac{dU}{T}}, \tag{2.26}$$

where $U \equiv U(\psi(T))$. If we write the energy in terms of the integral of a space density, and perform a Wick rotation from the real temperature axis to the imaginary one, in order to properly embed the time coordinate in the space-time metric, we obtain:

$$\mathcal{Z} \equiv \int \mathcal{D}\psi \, \mathrm{e}^{\mathrm{i} \int d^4x \, E(x)}. \tag{2.27}$$

Let's now define:

$$\mathcal{S} \equiv \int d^4x \, E(x). \tag{2.28}$$

Although it doesn't exactly look like, $\mathcal{S}$ is indeed the Lagrangian Action in the usual sense. The point is that the density $E(x)$ here is a pure kinetic energy term: $E(x) \equiv E_k$. In the definition of the action, we would like to see subtracted a potential term: $E(x) = E_k - \mathcal{V}$. However, the $\mathcal{V}$ term that normally appears in the usual definition of the action, is in this quantum gravity/ stringy framework a purely effective term, that accounts for the boundary contribution. Let's better explain this point. What one usually has in a quantum action in the Lagrangian formulation, is an integrand:

$$L = E_k - \mathcal{V}, \tag{2.29}$$

where E_k, the kinetic term, accounts for the propagation of the (massless) fields, and for their interactions. Were the fields to remain massless, this would be all the story. The reason why we usually need to introduce a potential, the $\mathcal{V}$ term, is that we want to account for masses and the vacuum energy (in other terms, the Higgs potential, and the (super)gravity potential). As we will discuss, in our scenario, a non-vanishing vacuum energy, as well as non-vanishing masses, are not produced, as in quantum field theory, through a Higgs mechanism. They are due to the finiteness of the extension of space-time, which reflects into mass gaps through so-called "stringy Higgs mechanisms", that don't require a Higgs field[14].

[13] There is no resistance opposed from something outside, simply because there is no outside, and, as we will also see in section 5.2, the expansion is not accelerated, $t = R$.

[14] In other terms, one can think that the Higgs fields are "frozen" at, or above, the string scale, that in our case will turn out to coincide with the Planck scale.

When we minimize 2.28 by a variation of fields in a finite space-time volume, we get a non-vanishing boundary term due to the non-vanishing of the fields at the horizon of space-time (moreover, we obtain also that energy is not conserved). In a framework in which space-time is considered of infinite extension, as in the traditional field theory, one mimics this term by introducing a potential term $\mathcal{V}$, which has to be introduced and adjusted "ad hoc", with parameters whose origin remains obscure. In particular, it remains completely unexplained why the cosmological constant, accounting for the "vacuum energy" of the Universe, as well as the other two contributions to the energy of the Universe, correspond to densities ρ_Λ, ρ_m, ρ_r, whose present values are of the order of the inverse square of the age of the Universe $\mathcal{T}$:

$$\rho \sim \frac{1}{\mathcal{T}^2} \, . \tag{2.30}$$

Were these "true" bulk densities, they should scale as the inverse of the space volume, $\sim 1/\mathcal{T}^3$. Here we understand why they instead scale not as volume densities but as surface densities: they are boundary terms, and as such they live on a hypersurface of dimension $d = dim[\text{space-time}] - 1$. We don't want to go more into details at this point of the discussion: along this work we will see how all this is precisely accounted for by string theory. For what concerns then the measure of integration, $\mathcal{D}\psi$, in 2.18 this indicates a sum over all string configurations. In the ordinary path integral, this is just a sum over all field configurations of one vacuum. But this is precisely the result of restricting $\{\psi\}$ to $\{\psi^{\min}(T)\}$.

Before concluding this section, we want just to remark that, in the light of this framework, the usual Lagrangian formulation 2.29 can be interpreted as follows: the range along which entropy can be varied to look for its minima is not the entire span of kinetic energy. The volume of the space at its disposal is "reduced" by the $\mathcal{V}$ term, which tells us that, at any temperature, we cannot go below the energy gap provided by the rest energy (masses). At infinity, $\mathcal{T} \to \infty$, the boundary term $\mathcal{V}$ vanishes. Notice that at the origin, $\mathcal{T} = 1$ in Planck units, the uncertainty relations tell us that both $\mathcal{V}$ and E_k must have a minimal value, $\mathcal{O}(1)$, which is also their maximal value, because the energy of the (observable) Universe is bound by the Schwarzschild black-hole relation. Therefore, $E_k - \mathcal{V} = 0$, and, as a consequence, also $S = - < \ln P >= 0$, in agreement with the fact that we expect a minimum of entropy, and just one, elementary configuration, with probability $P = 1$.

A first brief summary

Before to proceed with the analysis in this paper, we summarize the main points resulting from the discussion of this section.

- Finiteness of the speed of light, or, more in general, the existence of a maximal speed for the transfer of information, implies that the Universe we observe, i.e. our "causal region", is at any time of finite extension. Restricting the Universe to this region implies that space-time possesses a non-vanishing curvature. According to General Relativity, this is equivalent to an energy gap. The existence of a non-vanishing energy gap for any non-vanishing space-time fluctuation leads to an "uncertainty relation" corresponding to the usual Heisenberg's Principle. The theory able to explain what happens within this region must therefore be a quantum gravity theory.

- Another consequence of General Relativity is that, besides a minimal energy gap, this Universe possesses also a minimal length, the Planck length. Its quantum theory must possess a built-in symmetry under T-duality with respect to this length, which basically exchanges extended objects with "black holes", and enables the theory to deal with both kinds of objects. This leads to String Theory, as a basic ingredient of the description of the Universe.

- The system evolves in a random, probabilistic way, favouring a progress toward configurations with increasing space-time volume and entropy, realized by preferring steps of minimal increase of entropy, starting from the minimal entropy configuration at volume 1 in Planck units. All this is encoded in expression 2.18. This is the main result so far, the expression that collects all the information about the Universe, and constitutes an extension to quantum gravity of the Feynman's path integral. For convenience, we quote it also here:

$$\boxed{\mathcal{Z}_V = \int \mathcal{D}\psi \, \mathrm{e}^{-S(\psi)}} \tag{2.31}$$

  This can somehow be considered as the quantum version of the second law of thermodynamics. Instead of a "deterministic" progress toward the increase of entropy, we have here a probabilistic one. The "classical" thermodynamic law is approached only asymptotically: the probability for the system to become something closer and closer

to a classical thermodynamic system increases with its evolution. Tunnelling "backwards in time" is in principle not forbidden (as well as, with reference to the solution discussed in section 3, de-twisting of coordinates and transitions with non-minimal entropy increase [15]). However these effects are unfavoured, they lead to configurations much less probable than the "right" one. As space-time increases and the phase space becomes more complex, larger becomes the "gap" between more probable and less probable configurations. Therefore the Universe looks "mostly" as the one corresponding to the minimal entropy configuration, it is a superposition of states in which, as time goes by, the minimal, "classical solution" dominates the more and more [16]. As time goes by, the second law of thermodynamics is followed better and better.

In the next section, we will investigate the implications of equation 2.18 within the framework of string theory. To the effective geometry of space-time we will come back in section 5, where we will discuss how the geometry of a 3-sphere arises within the class of minimal entropy string configurations of the Universe.

[15] We will comment in section 11 also about the possible appearance of tachyons and non-critical string vacua.

[16] In a loose sense, although in a rather different theoretical framework, ideas like "considering all possible configurations", or the concept of looking for a class of vacua, better than a single, well defined string solution, turn out to be not so far from some aspects of the issues discussed in Refs. [6, 7, 8, 9, 10].

3 The non-perturbative solution

From now on, we are going to take a different approach with respect to the way we proceeded in the previous section. So far, we have investigated the properties of the Universe by considering the consequences of General Relativity on the geometry and physics of a space-time "defined" by the span of light rays at finite time, i.e. a space-time of finite extension. Expression 2.31 is the output of the previous section. Now this will be considered to be the starting point. It is even possible to leave aside the considerations through which we arrived to this proposal: now this is "the" formula we are going to investigate, within the framework of string theory. In principle, it contains all the informations about our Universe. However, in order to understand what in practice this means, we must "solve" it for the set of configurations that at best correspond to what we can measure and test. Any experimental output is given as the result of a series of measurements; this means that we always observe an average effect. Therefore, what we have to look for in our theory is the most favoured configuration at any volume/time, the $\psi^{\min}(\mathcal{T})$ of the previous section: this is the one which at best corresponds to the Universe as is experimentally observed; we will refer to it as to "the" vacuum. We stress however that $\psi^{\min}(\mathcal{T})$ is not a solution in a classical sense, but only the (largely) dominant configuration, that, as we will discuss in section 4.1, dominates the more and more at larger times/temperatures. The "infinite time" (zero temperature) limit is here the analogue of the classical limit of quantum mechanics, obtained by letting $1/\hbar \to \infty$. Indeed, from 2.31 one can see that configurations with entropy just a bit higher than the minimal one are not so suppressed with respect to the one of minimal entropy. However, we will find out that minimization of entropy leads to a separation of the string space into an internal one, frozen at the Planck scale, and what we indeed call space-time, free to expand. At large space-time volumes, changes in entropy can be due to a change in the configuration of the internal space, which leads to a different spectrum of the fundamental degrees of freedom, or to "macroscopic" changes in the distribution of these degrees of freedom (i.e. particles and fields) in the space-time. Changes of this second kind correspond to minor steps in the phase space. This means that, even if configurations with an entropy close to the minimal one contribute to the effective appearance of the Universe not much less than the configuration of minimal entropy, the "physics" they lead to is also very close. A small change of entropy leads to a small change in the physical configuration; it is in this sense that it is reasonable to consider an approach to the problem of minimization of entropy even in an approximated way.

In the following, we will restrict our analysis to supersymmetric string theory in critical dimension. We will comment on the most general case in section 11. In order to investigate the physical content of 2.18, we will use a "perturbative" approach. In ordinary quantum field theory one separates the time evolution into a free propagation and an interaction part. The actual configurations are inspected through the conceptual separation into a base of free states, eigenstates of the free Hamiltonian, which are exact solutions of the free theory. As long as the coupling of the interaction is small, the full solution can be considered as a small perturbation of the free propagation, and the perturbative approach makes sense. In our case, we have a truly non-perturbative string system, in which even space-time, intended as a whole, internal and external all together, is mixed up into something for which we don't even

know whether it is appropriate to talk of "geometry". In general, it will not be factorizable into an extended one, "the" space-time as we experience it, and an internal space: the one will be somehow embedded into the other one. Moreover, we can access to the whole theory only through "slices", the perturbative (string) constructions, to be treated as the patches, the "projections", which allow to shed light into the "patchwork", the whole theory. It is known, and we have also seen it in Ref. [11], that in many cases such an approach is possible. We will get information about the true vacuum through the investigation of the "patches", and a heavy use of string-string duality. As a basis of "free states" for the string configurations, we will use the class of orbifolds. In order to minimize entropy, we have in fact to freeze as many moduli as possible, and orbifolds are indeed singular spaces in which the highest amount of moduli of the "curved" space are "frozen": a smoother space is obtained from the orbifold point by un-freezing some moduli. Therefore, in this class of vacua the symmetry of the system is highly reduced, and, as a consequence, also entropy is. The orbifold point is however in some cases a point of enhanced symmetry. As we will discuss, this kind of degeneracies are removed by mass differentiations. The latter are driven by the moduli of space-time, related to non-twisted coordinates. This effect cannot be investigated at the orbifold point, where space-time comes factorized into a product of "orthogonal" coordinates: the effect of a non-trivial embedding of coordinates into the "true" space has then to be treated as a perturbation. These corrections are however expected to depend only on the non-twisted moduli, i.e. those of the "extended" space-time. If we generically indicate these moduli with R_i (basically they are "radii" of the extended, but nevertheless compact, space), we expect the relative corrections to be of order $\sim \mathcal{O}\left(\prod 1/R_i \times \log R_j\right)$. The role "S-duality" has in field theory is played here by T-duality: only if T-duality is broken the corrections at large radii are small, otherwise they would be of the same size as at small radii [17].

3.1 How to compute entropy in orbifolds?

Orbifolds are particular string constructions in which we have, at any energy level, full knowledge about the spectrum of the perturbative states. We are therefore able to write the partition function, the "one loop partition function", which in principle encodes all the information about the construction. This is defined as the sum over all states, weighted with the exponential of their energy. A relevant difference between the orbifold partition function and a thermodynamical partition function is that the first is by definition a sum at the one-loop order of expansion, the weight of the contribution of the states is an exponential in which energy appears squared instead of being at the first power, and at the place of the β parameter (basically the inverse of the temperature), we have the world-sheet parameter "τ", as dictated by the specific characters of two-dimensional conformal field theory, to be integrated over a specific domain. The details of the parameter τ (whether complex or real), and of the integration domain, depend on the type of string construction. A second, perhaps even more relevant, difference, is the fact that the contribution of states is also weighted with a sign, depending on their supersymmetry charge: the supersymmetric partners contribute

[17] After all, when seen from the point of view of string theory, "S-duality" is nothing else than a kind of T-duality.

with an opposite sign to the partition function, so that the latter vanishes for unbroken supersymmetry. Nevertheless, with the string orbifold partition function it is possible to perform one-loop computations of scattering amplitudes and threshold corrections. Therefore, within the restriction to the one-loop perturbative level, it works well as we expect from an ordinary partition function. The point is however rather subtle, and we must clarify several questions about it. Let's start by trying to define entropy for a string orbifold construction. According to its definition, entropy is given by:

$$S \equiv - \sum_i P_i \ln P_i \,. \tag{3.1}$$

Therefore, in order to be able to calculate the perturbative entropy, all what we need is to compute the probabilities P_i for all states i. With the help of string-string duality, we can then hope to learn also about the non-perturbative contributions to entropy. Let's for simplicity consider a closed string. The partition function is defined as:

$$\mathcal{Z} = \int_{\mathcal{F}} d\tau \sum (-1)^F q^{\mathrm{p}_{\mathrm{L}}^2} \bar{q}^{\mathrm{p}_{\mathrm{R}}^2} \,, \tag{3.2}$$

where $q \equiv \mathrm{e}^{2\pi \mathrm{i} \tau}$ ([18]). By considering that:

$$< A > = \sum_i A_i P_i = \sum_i < \emptyset | \left[|i> A_i < i| \right] | \emptyset > \,, \tag{3.3}$$

together with the fact that string amplitudes are computed, at one loop, as mean values of operators inserted in 3.2, we would be tempted to say that in the case of strings the probabilities are:

$$P_i \longrightarrow \int_{\mathcal{F}} d\tau (-1)^F q^{\mathrm{p}_{\mathrm{L}}^2(i)} \bar{q}^{\mathrm{p}_{\mathrm{R}}^2(i)} \,. \tag{3.4}$$

Instead of:

$$P_i = \frac{\mathrm{e}^{-\beta E_i}}{\mathcal{Z}} \tag{3.5}$$

we should have now:

$$P_i = \int (-1)^F \mathrm{e}^{-\operatorname{Im} \tau E_i^2} \,. \tag{3.6}$$

where it is intended that the integration over τ_2, the string torus complex structure, has already been performed, resulting in a delta function enforcing equal energies for the left and right movers. However, in attempting to force the interpretation of expression 3.6 as a probability we get into troubles. Apart from a problem of overall normalization, a first difficulty is due to the fact that, owing to the $(-1)^F$ factor, there are here states, the supersymmetric partners, which contribute negatively to the partition function, and therefore have, in the usual sense, a "negative probability". Our intuition suggests that

[18] As we said, the parameter τ admits in general several definitions, depending on whether we have an open or closed string. The partition function itself consists in general of a sum of several contributions, coming from closed and open string sectors. All these details don't affect the conclusions of the present discussion, that, for the sake of concreteness, we will focus to the case of closed strings.

both the $F = 1$ and $F = 0$ terms should give an equal contribution to the entropy of the system: the contributions of the two sectors should somehow add up. Normally, the total entropy of a system possessing a symmetry (i.e. larger phase space at disposal) is larger than the entropy of a system whose phase space is reduced.

Let's consider the way threshold corrections (or scattering amplitudes, as well) are computed. Usually, one is interested in the correction to the coupling of a certain term in the effective action; for instance, the $F_{\mu\nu}F^{\mu\nu}$ term, or the scattering amplitude of, say, gravitons. In order to compute it, one proceeds by inserting operators, or equivalently by switching on specific deformations through currents which do the same job; they namely lead to the insertion of the appropriate operators in the partition function. Although these operators are compatible with the symmetries of the string, among which is the underlying supersymmetry of the world-sheet conformal theory, as a matter of fact they break the target space supersymmetry. In order to isolate a single term, such as the bosonic $F_{\mu\nu}F^{\mu\nu}$ term, they *must* introduce an asymmetry, in this case between gauge bosons and their superpartners. For the purpose of performing the computation, supersymmetry is temporarily broken. It is restored afterwards, when the deformation is switched off. The correction to the supersymmetry related terms is then derived through a supersymmetry transformation. Therefore, as a matter of fact amplitudes are computed in a (virtually) non-supersymmetric vacuum, and then "pulled back", converted to a supersymmetric result by adding the missing terms, obtained through a supersymmetry transformation. In a situation of broken supersymmetry, the "vacuum" energy doesn't vanish anymore, and it is possible to talk about normalization of probabilities. The actual effective partition function on which these are computed is in practice:

$$\mathcal{Z}_{\rm eff.} \; = \; \int d\tau \sum_i (-1)^F \, \delta(F = 0/1) \, q^{\mathrm{p}_{\rm L}^{2(i)}} \, \bar{q}^{\mathrm{p}_{\rm R}^{2(i)}} \qquad \propto \; E_{\rm vac} \, , \tag{3.7}$$

The normalization of the partition function, the "vacuum energy" $E_{\rm vac}$, is then implicitly reabsorbed into a redefinition of the string mass scale, which does not appear when the amplitudes are converted to the duality-invariant Einstein's frame, where the string scale is substituted by the Planck scale (the latter is assumed to be the real, "physical" scale). It seems therefore that the natural environment in which the string partition function is used is the one of broken supersymmetry: even for supersymmetric vacua, what we do is to artificially break supersymmetry in order to switch on the term of interest, compute it in a phase of broken supersymmetry, and then come back to supersymmetry in order to complete the analysis of the effective theory. All this suggests that perhaps the correct environment in which the partition function has to be defined is the one of broken supersymmetry. In this way, it is not the phase of broken supersymmetry that must be seen as an artificial extrapolation, but rather the one of unbroken supersymmetry that must be viewed as a particular limit of the most general and natural situation of broken supersymmetry. By the way, we notice that, in a space-time with boundary, invariance under translations is broken. Since two supersymmetry transformations close on a space-time translation, we must conclude that in such a space also supersymmetry must be broken. Under this condition, the normalization of the partition function, no more vanishing, becomes possible.

Expression 3.7 is not yet satisfactory, it is affected by another problem: it counts all the

states for each energy level in the same way. This is correct as long as we don't reduce the symmetry of the vacuum. However, when we apply orbifold projections that reduce symmetries and supersymmetries, we introduce differentiations in the spectrum, which should reflect in a change in the entropy. The latter should be sensitive to the details of the new configuration. However, in orbifold constructions all degrees of freedom with the same energy/mass, although belonging to different (super)symmetry multiplets, seem to be blindly summed up. To be more precise, the orbifold partition function is defined as a sum over sectors with distinguished boundary conditions. Mean values of effective couplings and threshold corrections are computed through the insertion of operators, whose role is that of switching on deformations which select some of these sectors. Therefore, it is not true that the orbifold partition function is so "blind". However, it is true that, without these deformations of the vacuum, once resummed states with the same mass are indistinguishable. Indeed, in the moduli space of string theory orbifolds represent somehow points of "enhanced symmetry". As we will discuss in section 6, in the "real" minimal entropy vacuum this degeneracy is lifted by slight mass differences, which distinguish all the light states, namely those that at the orbifold point look all massless. The orbifold sits at a "limit", or corner, in the moduli space, in which these mass deformations are pushed to their extremal values: not only the partners corresponding to the broken supersymmetries in general simply disappear, projected out of the spectrum, but also the gauge bosons of the broken symmetries, namely those that were relating states of different orbifold sectors, are sent to infinite mass, while at the same time the states rotated by these symmetries are T-dually sent all to zero mass. In the resummed partition function they appear therefore all on the same footing, although in the "physical" configuration they are not. In the purpose of identifying the minimal entropy configuration within the class of orbifold vacua, we must therefore take into account also this "orbifold singularity".

Let's start by considering the maximal super-symmetric unprojected string vacuum ($\mathcal{N}_4 = 8$ [19]), and assume that, at least perturbatively, the "true" partition function is given by:

$$\mathcal{Z} \, = \, \int_{\mathcal{F}} d\tau \sum_i q^{\mathrm{p}_{\mathrm{L}}^{2(i)}} \bar{q}^{\mathrm{p}_{\mathrm{R}}^{2(i)}} \, . \tag{3.8}$$

This expression is similar to the usual "one loop partition function", with the difference however that here we are not subtracting the contribution of the superpartners, but adding them. As compared to the usual statistical-thermodynamical approach, the probabilities are:

$$P_i \, = \, \frac{\mathrm{e}^{-\beta E_i}}{Z(\beta)} \; \longleftrightarrow \; \frac{\int_{\mathrm{t}} \mathrm{e}^{-\mathrm{t} E_i^2}}{\int_{\mathrm{t}} Z(\mathrm{t})} \, , \tag{3.9}$$

[19] Along this work, we will use the notation $\mathcal{N}_d$ in order to indicate the number of space-time supersymmetries in d dimensions. In particular, for convenience most of the time we will count supersymmetries from the point of view of four space-time dimensions, regardless of the real number of compact coordinates. Intrinsically, $\mathcal{N}_4 = 8$ means nothing else than 32 supercharges. Incidentally, we recall that in our case all coordinates are compact. We will see that in our case the distinction between extended and compact space is established at the level of the scale of the coordinate size.

where "t" indicates the real integration parameter, after a possible identification of the left with the right-moving momenta. In the case of the closed string, it is the imaginary part of the world-sheet parameter: t $= \tau_1$. When we apply an orbifold projection, the amount of symmetry of the system is reduced. Let's consider the case of Z_2 orbifold operations, the case that will mostly interest us in the following. Among this class of projections, let's for the moment consider the case of a non-freely acting orbifold. In this case, a projection does not reduce the total number of states: the lost states are recovered at the fixed points, as twisted states. What however changes after the projection is the "nature", or identification of these states: the new states don't belong to a symmetry multiplet together with the unprojected states. From a statistical point of view, this is like having reduced by one half the states of the initial system, but having paired them with a new system, containing as many states as those missing in the first one. We can figure out the situation by representing the initial physical system as a gas of identical particles. Indeed, a string vacuum (and an orbifold in particular) can in fact also be thought as a gas, in which different particles act as sources for the singularities of the space (from a gravitational point of view, particles *are* singularities of space-time, being sources of gravitational field, singular points of the curvature). The orbifold projection introduces a "distinction mark" on half of the states, it labels them in another way. The situation is depicted in figure 6. The resulting orbifold follows now the laws of composite systems: the probability of the various configurations of the new string "vacuum" are given by the product of the probability of the unprojected part ("A" in figure 6) times the probability of the twisted part ("B" in figure 6):

$$P_{A+B} \; = \; P_A \times P_B \, . \tag{3.10}$$

If we want to derive these probabilities from the partition function as defined in 3.8, we must first separate it into the distinguished contributions of A and B, $\mathcal{Z}_A$ and $\mathcal{Z}_B$, extract the terms corresponding to the wanted energy level, and normalize them to the new, composite partition function. As we said, this is what is implicitly assumed also in any "threshold correction" computation ([12, 13, 14, 15, 16, 17, 18, 19, 20, 21]), although, owing to the lack of meaning of the overall normalization in the "supersymmetric definition" of the partition function 3.4, this issue is completely irrelevant and therefore is never addressed [20].

Another problem for the computation of entropy, is that in general we don't know the full partition function of a string construction. In particular, in the case of reduced supersymmetry non-perturbative sectors appear, e.g. corresponding to "D-branes" states. Even in the case of orbifold constructions, there is no perturbative approach allowing to explicitly see (i.e., construct the vertex operators for) all the states of the theory at once: we can know something about the full spectrum only by making heavy use of string-string duality. Fortunately, for our purpose we are not interested in an exact computation of entropy, but

[20] The fact that the system is a product of systems and true probabilities are factorized is automatically taken into account in threshold corrections by the appropriate operator insertion, that doesn't work "blindly" but picks just the subset of terms of relevance, factorizing out those corresponding to the other factors of the symmetry group. Namely, the symmetry group is in general a product $G = G_1 \times \cdots \times G_n$; the operator under consideration picks just the contribution of one of these factors (for instance, a factor of the gauge group if we are dealing with the correction to a $F_{\mu\nu}F^{\mu\nu}$ term). Up to an overall renormalization, this effectively allows to work with the string partition function as if it was a "true" partition function.

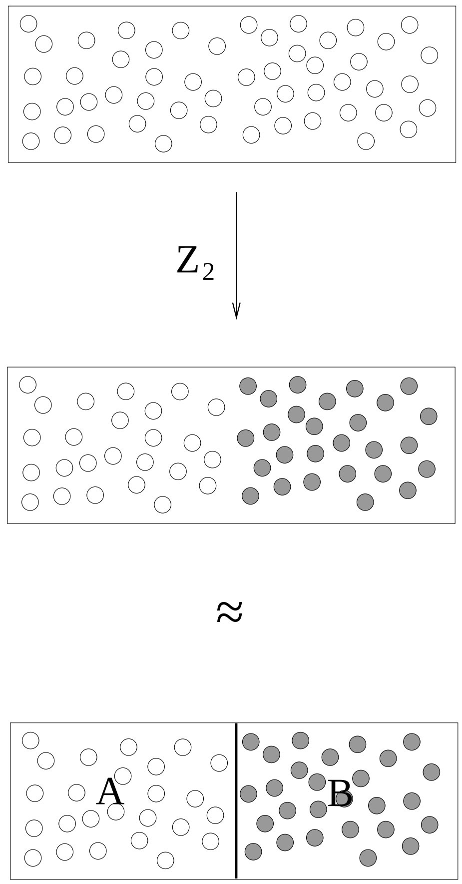
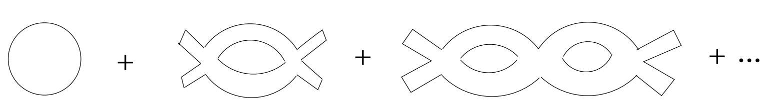


Figure 6: the effect of an orbifold projection is that of reducing the symmetry of the system, separating the spectrum in two sectors “confined” to different parts of the phase space.

just in its minimization. We can therefore proceed by assuming that, if entropy is minimized in all the "slices" of the theory we can perturbatively construct, then it is also minimized in its whole. Certainly, the concept of dual constructions is somehow similar to the idea of "projection" on a subspace. However, the dual perturbative slices don't correspond in general to a simple "truncation" in the space of the parameters, but are obtained through some limiting procedure, according to which some parameters are pushed to some corner in the moduli space, i.e. outside of the actual configuration we want to investigate. The concept of "tangent space" seems to be more suitable in order to represent what we are doing when we look at dual constructions: any such perturbative dual constitutes an approximation of the real thing, obtained by "linearizing" around some fixed value. Our problem is then to understand if, whenever the entropy of the vacuum "A" is lower than the entropy of vacuum "B" in all the perturbative slices, the ordering $S(A) < S(B)$ is preserved at the real point of the vacuum. This seems a reasonable assumption, because in passing from the non-perturbative vacuum to one of its perturbative slices involves a transformation of a modulus from a fixed (in general Planck-size) modulus/coordinate, to a small/large, i.e. zero/infinite value. Basically, we expect therefore a correction to entropy to be smaller than those driven by the extended space-time coordinates; in any case, smaller than the gap between entropies of vacua differing by a change in the orbifold configuration, which always involves different assets of finite (in general Planck-size) moduli/coordinates. If we ideally run back the various slices from the limit value at some corners in the moduli space, to the "bulk", we can expect to obtain a series of "patches" whose union "covers" the full, non-perturbative vacuum. Since entropy is an additive function, having minimized it on all the parts ensures us that we have minimized it in its whole.

For this purpose, indeed the one-loop partition function of the various perturbative slices turns out to be "the" partition function we must use. In principle, one could think that considering also higher perturbative orders can only improve the analysis. This is not true. The reason is that higher order terms are pieces of an expansion around a coupling of the perturbative vacuum, and as such give only incomplete information about pieces which are better obtained, in a non-perturbative way, from dual configurations. Let's consider a full, non-perturbative vacuum, characterized by n parameters (moduli) $X_i, \ldots, X_\ell$. Different perturbative dual realizations will correspond to series built around limit values of some of these moduli. Let's consider the simple case of a "slice" obtained as a perturbation around X_ℓ (think for instance at the heterotic string, and imagine that X_ℓ corresponds to the heterotic dilaton). The perturbative string expansion consists of a series of terms, given by world-sheet geometries ordered according to their "genus" g, corresponding to powers of the coupling field. For instance, a "four point" diagram is computed as:



g = 0 g = 1 g = 2 (3.11)

The terms $g = n$ correspond to the power $[e^{2\phi}]^{n-1}$ in the expansion, where ϕ is the dilaton field: $\phi \leftrightarrow X_\ell$. Only at genus one the mean value of the operator (in this case a four point vertex) can be written without ambiguity in terms of insertion into a partition function, because the $g = 1$ is the only term which does not depend on the coupling field:

$$< A >= \sum_i \int_{\mathcal{F}} d\tau A_i q^{\mathrm{p}_{\mathrm{L}}^{2(i)}} \bar{q}^{\mathrm{p}_{\mathrm{R}}^{2(i)}} , \tag{3.12}$$

The energies (masses) of the states contributing to the sum don't show any dependence on the coupling modulus X_ℓ: any such term would be non-perturbative. Indeed, as we will discuss later on in this work, orbifold constructions correspond to a linearization of the string space, and correspond to working "on the tangent space"; we have therefore to do with a kind of logarithmic representation: products of coordinates are mapped to sums. This allows to see non-vanishing masses also in a decompactification limit of the coupling: on the tangent space, the contribution of the coupling to these mass terms, instead of multiplying those of the perturbative moduli, sums up, and decouples without suppressing them. If the coordinates $X_i, \ldots, X_k$ are perturbative, and X_ℓ is non-perturbative, we can write a partition function only for the states whose mass is given in terms of windings and momenta of $R_i, \ldots, R_k$, whereas masses that go like $m \sim m/R_\ell + nR_\ell$ are completely hidden. The other terms (genus zero and higher than one) of the series illustrated in 3.11 are precisely part of an expansion around the coupling field, and don't contain a correct information about the mass of states: this would show up only non-perturbatively. These terms must therefore be investigated through dual "slices" of the theory, where they are perturbative. In the following, we focus our attention on one of these slices. Let's assume that we know the entropy of the string vacuum before the orbifold projection. We want to see how this quantity changes if we apply a Z_2 projection. Intuitively, it is clear that, since a projection reduces the amount of symmetry, it reduces also the volume of the phase space at disposal for the degrees of freedom. As a consequence of the increased concentration of the probability distribution, also entropy should be reduced. In order to see how precisely things work, we must first consider that a Z_2 projection divides the initial system in two parts, as is clear from the partition function:

$$\mathcal{Z} \xrightarrow{Z_2} \frac{1}{2}\left(\mathcal{Z}\begin{bmatrix}0\\0\end{bmatrix} + \mathcal{Z}\begin{bmatrix}0\\1\end{bmatrix}\right) + \frac{1}{2}\left(\mathcal{Z}\begin{bmatrix}1\\0\end{bmatrix} + \mathcal{Z}\begin{bmatrix}1\\1\end{bmatrix}\right) . \tag{3.13}$$

This corresponds to the process illustrated in figure 6. In order to understand in which direction the variation of entropy goes, we can view the process of separation of the phase-space into two sectors as the opposite of the adiabatic expansion illustrated in figure 7. In this example, the system is constituted by the sum of two subsystems of particles, that for convenience we have labelled with a different colour. Intrinsically, they are on the other hand indistinguishable; when they are together, such as on the l.h.s. of the figure, all particles are on the same footing, having at disposal the full phase space. Our system is the sum of the two systems on the r.h.s., and entropy is the sum of the two corresponding entropies. It is then clear that, owing to the process of "separation" induced by the orbifold operation, entropy decreases.

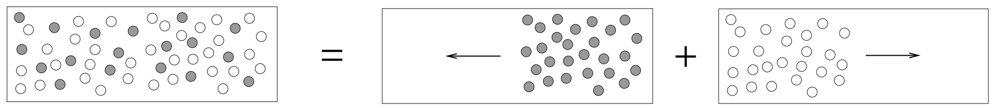


Figure 7: the entropy decrease due to an orbifold projection can be understood if we represent the system as the sum of two parts each one consisting of a gas of particles. The orbifold process is the reverse of the expansion of each of the two gases from half of the phase space until they occupy the entire phase space.

We can compute the amount of this reduction. Although represented as a "gas", the system does not follow the laws of gas thermodynamics: the particles don't have a "temperature" related to a "kinetic energy". In order to understand what happens to entropy, we must only think in terms of phase space and probability distributions. We can consider the whole system as ideally divided into two subsystems, A and B. The probability of each half system is 1/2 of the total probability:

$$P_A \,=\, P_B \,=\frac{1}{2}\,, \tag{3.14}$$

When this distinction is just virtual, as it is before the orbifold projection, the total probability of the system is $1 = 1/2 + 1/2$. As illustrated in figure 6, the Z_2 projection acts like the insertion of a wall between A and B, that prevents a mixing of the two parts. In this case, the probability of the configuration "A B" is the product of the two probabilities, because the phase space too has become a product:

$$P_{AB} \,=\, P_A \times P_B \,=\frac{1}{2}\times\frac{1}{2}\,=\,\frac{1}{4}\,. \tag{3.15}$$

Entropy of the initial system S_{A+B} is given by the sum $S_A + S_B$:

$$S_{A+B} \,=\, 2\,S_A \,=\, \ln 2\,. \tag{3.16}$$

After the projection, we have instead:

$$S_{AB} \,=\, -\,P_A P_B \ln P_A P_B \,=\, \frac{1}{2}\ln 2\,. \tag{3.17}$$

Entropy is therefore reduced by half.

The amount of reduction is in agreement with the behaviour we expect for entropy in this string framework. In section 4 we will in fact see more in detail that the entropy of the Universe scales as its radius, or age, to the second power: $S \approx \mathcal{T}^2$, and indeed, one can see that such a Z_2 orbifold projection effectively halves the volume of the target space, so that $\mathcal{T} \to \mathcal{T}/\sqrt{2}$. For simplicity, consider the case of a four dimensional construction. Indeed, our situation corresponds to a compactification to zero-dimensions, being all string coordinates compact. However, in this four dimensional case there are already results available in the literature. Generalizing to the case of a compactification to lower dimensions is trivial. Let's then consider a toroidal compactification to four dimensions. After a Z_2 orbifold projection, the compact space becomes the product of a twisted four dimensional part, equivalent to a K3 manifold, and a two-dimensional, untwisted torus. Usual investigation of threshold corrections for $\mathcal{N}_4 = 4$, Z_2 orbifolds in the type II string, or $\mathcal{N}_4 = 2$ orbifolds in the heterotic string, in particular the corrections to effective R^2 (and $F_{\mu\nu}F^{\mu\nu}$) couplings [21], show that they explicitly depend on the moduli of this torus. In the limit of large volume, they scale linearly with the volume:

$$\frac{1}{g^2_{\mathcal{N}}} \sim T \, , \tag{3.18}$$

where T is the modulus corresponding to the torus volume form. In the type II string, we can explicitly follow the fate of this modulus even after a further orbifold projection. In this case, we have:

$$\frac{1}{g^2_{\mathcal{N}/2}} \sim \frac{T}{2} \, . \tag{3.19}$$

The factor 1/2 is introduced by the new projection. It seems therefore that, owing to the projection, the volume of the untwisted torus has been halved:

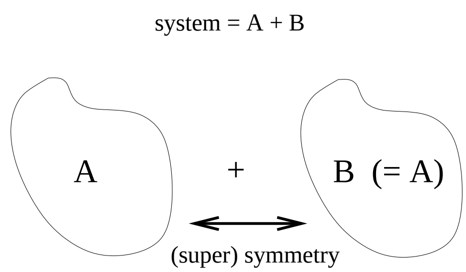
$$T \to \frac{T}{2} \, . \tag{3.20}$$

This phenomenon can be observed also in the heterotic and type I duals [22, 16, 11], where this modulus is identified with the coupling field S. After the projection, $S \to S/2$. It may look strange that an orbifold operation reduces the volume of a part of the space which has not been involved in the projection. The point is that what appears in the corrections is a function of the volume of the untwisted (two-) torus times a function of the volume of the twisted torus: $\sim V_1 V_2$. It is somehow arbitrary to refer the projection factor $\frac{1}{2}$ to V_1 or to V_2. The correct interpretation is that the projection halves the volume of the target space by halving the length of one of its coordinates. According to the scaling of entropy, we expect therefore that this too is reduced by a factor 2.

To summarize, under orbifold operation the space volume is reduced to one half, while the number of degrees of freedom remains unchanged: half of them are projected out, but an equal number appears now in the twisted sector. However, the new states are independent, distinguished from the lost ones: they don't belong anymore to a multiplet together with

[21] See for instance Refs. [12, 22, 23, 24, 25, 13, 14, 18, 26, 16, 17, 27, 28, 29], and, for the type I string, [19, 20, 21, 30, 31].

those which have not been projected out. Before the projection, we had a system replicated by (super)symmetry:



After the projection, we have in volume A half of the previous states *plus* the new states from the twisted sector, and no replica-volume B. The volume has been halved, but the states are in the same number as the initial ones. From an effective point of view, what we have done is indeed equivalent to a volume contraction at fixed number of states.

We consider now the case of an orbifold in which the Z_2 acts freely, by shifting instead of twisting the coordinates. Besides this, there exist another class of orbifold constructions, which are "in between". They are the so called "semi-freely" acting orbifolds. They are obtained through projections consisting of twists and shifts.

Let's first discuss the case of a true, pure freely acting orbifold. A shift is an operation that, instead of completely projecting out of the spectrum the odd states, just lifts their mass by assigning them a non-vanishing momentum and/or winding number along the shifted coordinate. Intuitively, also a shift reduces entropy, because it reduces the amount of symmetry by introducing differentiations in the spectrum. Differently from the case of non freely acting orbifolds, there are however no new sectors (twisted sectors) generated by the projection: states are here distinguished through a mass gap. In order to understand the change in entropy due to this kind of projection, we don't have therefore to proceed by considering the system as a complex one, consisting of various sectors whose probabilities must be multiplied, as in 3.10. Here we can focus our attention to the transformations operated on the terms of the initial partition function. The breaking of symmetry introduced by the freely acting projection lifts the mass of the odd states. The consequence is that the probability of the lifted states is lower than the one of the non-lifted ones:

$$P_{\text{lift}} \sim \frac{\int 1}{\mathcal{Z}} \to \frac{\int \mathrm{e}^{-\tau E^2}}{\mathcal{Z}^{\mathrm{s}}}, \tag{3.21}$$

where by $\mathcal{Z}$ we indicate the original partition function, by $\mathcal{Z}^{\mathrm{s}}$ the partition function with the orbifold shift. We have that:

$$\mathcal{Z}^{\mathrm{s}} \approx \mathcal{Z} \times \frac{1}{2}\left(1 + \mathcal{O}\left(\frac{\int \mathrm{e}^{-\tau E^2}}{\int 1}\right)\right). \tag{3.22}$$

For the non-lifted states, after the projection the probability becomes:

$$P \sim \frac{\int 1}{\mathcal{Z}} \xrightarrow{Z_2^{\rm s}} \frac{\int 1}{\mathcal{Z}^{\rm s}} \,. \tag{3.23}$$

For the lifted states, the ratio of probabilities after/before the orbifold shift is:

$$P^{\rm s}/P \;\lesssim\; \frac{\int {\rm e}^{-\tau E^2}}{\int 1} \times \frac{\mathcal{Z}}{\mathcal{Z}^{\rm s}} \,. \tag{3.24}$$

By using 3.22, this can be rewritten as:

$$P^{\rm s}/P \;\lesssim\; \frac{\int {\rm e}^{-\tau E^2}}{\int 1} \times 2 \left(1 - \mathcal{O} \left(\frac{\int {\rm e}^{-\tau E^2}}{\int 1} \right) \right) \,. \tag{3.25}$$

In order to see that it is always < 1, just rewrite it as:

$$P^{\rm s}/P \;\lesssim\; 2\,x(1-x) \,, \qquad 0 \,<\, x \,<\, 1 \,, \tag{3.26}$$

where $x \equiv \int {\rm e}^{-\tau E^2} / \int 1$. This shows that the distribution has become more picked around the non-lifted states. The probability distribution is more concentrated, entropy has decreased. On the other hand, the reduction of entropy due to the shift in a freely acting orbifold is smaller than the reduction produced by an equivalent non-freely acting projection. In order to see it, consider again the system as a "gas" that gets divided into two parts, as we did for a non-freely acting orbifold, illustrated in figure 6. The difference is that in the case of free orbifold action the "B" states don't have the same mass as the "A" states. As we have seen, their probabilities are lower. Since anyway the sum of probabilities must be 1, we can write:

$$P_A^{\rm free} \,=\, P_A \,+\, \varepsilon \qquad\qquad P_B^{\rm free} \,=\, P_A \,-\, \varepsilon \,, \tag{3.27}$$

and compute entropy as in 3.17:

$$\begin{aligned} S_{AB}^{\rm free} \;&=\; -(P_A+\varepsilon)(P_B-\varepsilon)\,\ln\left[(P_A+\varepsilon)(P_B-\varepsilon)\right] \\ &=\; -(P^2-\varepsilon^2)\,\ln(P^2-\varepsilon^2) \,, \end{aligned} \tag{3.28}$$

where we have substituted $P_A \,=\, P_B \,\equiv\, P$. We want now to compare 3.28 with 3.17, $S_{AB} \,=\, -P^2 \ln P^2$, to show that now the decrease of entropy due to the orbifold projection is lower. The result we are looking for is due to the properties of the function $S \,=\, -x \ln x$ for $0 < x < 1$, which increases as $x = P^2$ decreases: this implies that $S_{AB}^{\rm free} \,\equiv\, S(x-\varepsilon^2) \,>\, S_{AB} \,=\, S(x)$.

Semi-freely acting orbifolds are obtained with a Z_2 projection that acts only in part freely. They have therefore some twisted states, although not as many as a non-freely acting orbifold. They are therefore an intermediate step between freely and non-freely acting ones. It is not difficult to realize that also the decrease of entropy due to this kind of projections

is something intermediate between the higher one of a non-freely acting and the lower one of a freely-acting orbifold.

Coming back to our main problem, namely the search for the orbifold vacuum that minimizes 2.31, after this discussion it is clear that the most favoured configuration will be the one with the maximal amount of non-freely acting orbifold projections, followed in turn by semi-freely and completely freely acting ones. Moreover, Z_2 orbifolds will be the most favoured ones, because they mod-out the space by the group with the smallest volume among all the orbifold operations. A product of Z_2 twist/shifts allows therefore to achieve a configuration which, having a smaller surviving symmetry group, is more concentrated than those obtained through any other product of orbifold operations. Entropy will therefore be the minimal we can obtain with orbifold operations.

3.1.1 A note on the gravitational collapse

In section 3.1, in order to understand how things work for string orbifold configurations, we have compared the geometric space to a gas of particles. This pictorial representation was based on the consideration that particles are sources for singularity of the space-time geometry, as much as the fixed points of an orbifold are for the internal string space. On the other hand, we have also previously seen that the physical evolution proceeds from lower to higher entropy configurations. A possible remark could then be that the above pictorial representation becomes problematic precisely in the case we want to consider geometry of space-time, and therefore gravitation: according to our arguments, evolving toward an increase of entropy basically means evolving toward less singular, more spread out configurations. Once gravity is included in the game, this seems to be contradicted by the case of a gravitational collapse, a physical process in which space-time becomes more singular. Indeed, gravity is a reversible force: we expect therefore that during a process driven by the gravitational force, entropy does not change. Certainly, according to the discussion presented in this chapter, the evolution toward a more singular space-time configuration should lead to a decrease of entropy; and indeed, in itself, it does. However, the energies of particles involved in the process (for instance, kinetic energies), in short, the temperatures, increase. Once all this is taken into account, it is easy to realize that this leads to an increase of entropy, as a consequence of the increased spreading out of probabilities in the phase space. In the case of a reversible process, this increase compensates the decrease due to a higher singularity of the space-time configuration. In a non-reversible process, it exceeds it. In general, a physical system certainly evolves toward higher, or at least not lower, entropy configurations. These however don't necessarily correspond to more "spread out" layouts of the space-time geometry. The real "geometry" to be considered in the balance is that of the full phase space.

3.2 *Investigating orbifolds through string-string duality*

Investigating the non-perturbative properties of a string vacuum by comparing dual constructions is neither an easy task, nor a straightforward one. In general, at a generic point in moduli space the full set of dual constructions, enabling to "cover" the full content, is not known. Some progress on this knowledge has been done in the case of supersymmetric vacua with extended supersymmetry, where it is in general possible to identify a subset of the spectrum made "stable" by the properties of supersymmetry. The case of orbifolds turns out to be particularly suited for the investigation of non-perturbative string-string dualities. In this case it is possible to make a non-trivial comparison of the renormalization of terms that receive contributions only from the so called BPS states, and this not just on the ground of the properties of supersymmetry, but through the computation of true string contributions. Fortunately, Z_2 orbifolds, the case of our interest, are the easiest and therefore more investigated constructions [22]. Indeed, through the analysis of these constructions, it is possible to get an insight into the properties which are typical of string theory in itself: most of the investigations performed at other points in the moduli space must in fact rely on geometrical properties of smooth surfaces, and their singularities. Although for some respects rather powerful, these techniques don't allow to capture the presence of states related to non-geometrical singularities, or even fail in general for the simple reason that, owing to T-duality, the full string space simply cannot be reduced to a geometrical one[23]. In the following, we will repeat the steps of entropy reduction already discussed in Ref. [2]. As we said, in the purpose of minimizing entropy the most favoured Z_2 operations are non-freely acting orbifold twists. Therefore, the string vacuum we are looking for will have the maximal possible amount of such operations. Our starting point will be a maximally supersymmetric string vacuum with flat background: in our approach, the curvature of space-time will come out as an output, it is not an external input of the theory. The constraints of two-dimensional conformal field theory impose that Z_2 orbifold twists must act on groups of four coordinates at once. In any string construction, there is room for a maximum of 3 such operations, one of which is however redundant, in that it leads, once combined with the other ones, to the re-introduction in the twisted sectors of the states projected out. Therefore, we can say that only a maximum of two independent Z_2 twist act effectively. However, the amount of supersymmetry surviving to these projections, as well as the amount of initial supersymmetry, is different, depending on whether we start with heterotic, type I, or type II strings. This means that in any construction not all the projections acting on the theory are visible. Indeed, one of them is always non-perturbative. The reason is that, by definition, a perturbative construction is an expansion around the zero value of a parameter, the coupling of the theory, which is itself a coordinate in the whole theory. An orbifold operation acting on this coordinate is forcedly non-perturbative. A first investigation of a non-perturbative orbifold, which produces the heterotic string, has been carried out in [43, 44]. The question about how many such "hidden coordinates" indeed exist in the underlying, full theory, remains however still unanswered. All what we know is

[22] Besides the works already cited at the pages 42 and 47, see also for instance Refs. [32, 11, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42].

[23] For examples, see for instance Ref. [11].

that there exist no flat supersymmetric vacua in dimensions higher than $d = 11$.

Our aim is now to derive the structure of the minimal entropy vacuum, independently on the size of the target space. However, our basic hypothesis is that space-time is always compact. A perturbative construction is on the other hand built around a small/vanishing value of a coupling, which, in the string language, is a coordinate. Orbifolds are therefore built at decompactification limits. Dual constructions correspond in general to different decompactifications, i.e. of different coordinates. Essential for our work is that such a decompactification must be possible; for this to make sense, the involved coordinate(s) must not be twisted. Even in the case it is shifted, under decompactification the scenario gets "trivialized" and loose information about the construction. This problem does not exist for the first steps of entropy/symmetry reduction. It becomes however relevant when we reach the maximum of orbifold operations. As we will see, for the minimal entropy configuration, indeed a decompactification limit is a trivialization of the theory, under which some physical content is lost: a perturbative treatment of the degrees of freedom does not correspond in this case anymore to a simple "limit" of the theory, but to a phase which can be obtained only through a "logarithmic mapping" of the coordinates of the physical vacuum. This will be a key point of the entire discussion of masses and supersymmetry breaking. In this section we are not concerned with this problem; in the purpose of understanding what is the singularity structure, we proceed, as in Ref. [2], by first "counting" the Z_2 operations in the various decompactification limits, as long as these can be taken. By the way, we remark that it is precisely thanks to this limiting procedure that it is possible to construct supersymmetric string constructions: as we already pointed out, in a fully compact space-time supersymmetry is always broken.

In the following we will often make use of the language of string compactifications to four dimensions, especially for what matters our reference to the moduli of the string orbifolds. This will turn out to be justified "a posteriori": we will see that indeed the final configuration is the one of a string space with all but four coordinates twisted and therefore "frozen". Only four coordinates remain un-twisted and free to expand, while all the others remain stuck at the string/Planck scale. Massless degrees of freedom move along these and expand the horizon of space-time at the speed of light. Although not infinitely extended, this "large" space is what in our scenario corresponds to the ordinary space-time. The language of orbifold constructions in four dimensions is therefore just an approximation, that works particularly well at large times. Only at a second stage, we will also discuss how and where this picture must be corrected in order to account also for compactness of the space-time coordinates. Although somehow an abuse of language, this approximation allows us to take and use with little changes many things already available in the literature. In particular, for several preliminary results and a rediscussion of the previous literature, the reader is referred to Ref. [11].

Let's see what are in practice the steps of increasing singularity/decreasing entropy we encounter when approaching the most singular configuration. Starting from the M-theory configuration with 32 supercharges, we come, through orbifold projections, to 16 supercharges and a gauge group of rank 16. Further orbifolding leads then to 8 supercharges ($\mathcal{N}_4 = 2$) and introduces for the first time non-trivial matter states (hypermultiplets). As we

have seen in [11] through an analysis of all the three dual string realizations of this vacuum (type II, type I and heterotic), this orbifold possesses three gauge sectors with maximal gauge group of rank 16 in each. The matter states of interest for us are hypermultiplets in bi-fundamental representations: these are in fact those which at the end will describe leptons and quarks (all the others are eventually projected out). As discussed in [11], in the simplest formulation the theory has 256 such degrees of freedom. The less entropic configuration is however the one in which, owing to the action of further Z_2 shifts, the rank is reduced to 4 in each of the three sectors. These operations, acting as rank-reducing projections, have been extensively discussed in [45, 26, 27, 11]. The presence of massless matter is in this case still such that the gauge beta functions vanish. In this case, the number of bi-charged matter states is also reduced to $4 \times 4 = 16$. These states are indeed the twisted states associated to the fixed points of the projection that reduces the amount of supersymmetry from 16 to 8 supercharges.

Let's consider the situation as seen from the type II side. We indicate the string coordinates as $\{x_0, \ldots, x_9\}$, and consider $\{x_0, x_9\}$ the two longitudinal degrees freedom of the light-cone gauge. The transverse coordinates are $\{x_1, \ldots, x_8\}$. Here all the projections appear as left-right symmetric. The identification of the degrees of freedom, via string-string duality, on the type I and heterotic side depends much on the role we decide to assign to the coordinates, as we will see in a moment. By convention, we choose the first Z_2 to twist $\{x_5, x_6, x_7, x_8\}$:

$$Z_2^{(1)}: \quad (x_5, x_6, x_7, x_8) \rightarrow (-x_5, -x_6, -x_7, -x_8)\,, \tag{3.29}$$

and the second Z_2 to twist $\{x_3, x_4, x_5, x_6\}$:

$$Z_2^{(2)}: \quad (x_3, x_4, x_5, x_6) \rightarrow (-x_3, -x_4, -x_5, -x_6)\,. \tag{3.30}$$

These two projections induce a third one: $Z_2^{(1,2)} \equiv Z_2^{(1)} \times Z_2^{(2)}$, that twists $\{x_3, x_4, x_7, x_8\}$:

$$Z_2^{(1,2)}: \quad (x_3, x_4, x_7, x_8) \rightarrow (-x_3, -x_4, -x_7, -x_8)\,. \tag{3.31}$$

Altogether, they reduce supersymmetry from $\mathcal{N}_4 = 8$ to $\mathcal{N}_4 = 2$, generating 3 twisted sectors. Depending on whether we consider the type IIA or IIB construction, the twisted sectors give rise either to matter states (hyper-multiplets) or to gauge bosons (vector-multiplets). As we discussed in Ref. [11], comparison with the heterotic and type I duals shows that the underlying theory must be considered as the union of the two realizations: owing to the lack of a representation of vertex operators at once perturbative for all of them, for technical reasons no one of the constructions is able to explicitly show the full content of this vacuum. The matter (and gauge) content in these sectors is then reduced by six Z_2 shifts acting, two by two, by pairing each of the three twists of above with a shift along one of the two coordinates of the set $\{x_1, \ldots, x_8\}$ which are not twisted. Each shift reduces the number of fixed points of a Z_2 twist by one-half; two shifts reduce therefore the matter states of a twisted sector from 16 to 4. Altogether we have then, besides the $\mathcal{N}_4 = 2$ gravity supermultiplet, three twisted sectors giving rise each one to 4 matter multiplets (and a rank 4 gauge group). On the type I side, these three sectors appear as two perturbative D-brane sectors, D9 and D5, while the third is non-perturbative. On the heterotic side, two sectors are non-perturbative. As it

can be seen by investigating duality with the type I and heterotic string, the matter states from the twisted sectors are actually bi-charged (see Refs. [46, 47], and [11]), something that cannot be explicitly observed, the charges being entirely non-perturbative from the type II point of view. The moduli $T^{(1)}$, $T^{(2)}$, $T^{(3)}$ of the type II realization, associated respectively to the volume form of each one of the three tori $\{x_3, x_4\}$, $\{x_5, x_6\}$, $\{x_7, x_8\}$, are indeed "coupling moduli", and correspond to the moduli "S", "T", "U" of the theory. On the heterotic side, S is the field whose imaginary part parametrizes the string coupling: $\mathrm{Im}\, S = \mathrm{e}^{-2\phi}$. It is therefore the coupling of the sector that contains the gravity fields. T and U are perturbative moduli, and correspond to the couplings of the two non-perturbative sectors. On the type I side, on the other hand, two of them are non-perturbative, coupling moduli, respectively of the D9 and D5 branes, while only one of them is a perturbative modulus, corresponding to the coupling of a non-perturbative sector [46, 48, 49, 35]. Owing to the artifacts of the linearization of the string space provided by the orbifold construction, gravity appears to be on a different footing on each of these three dual constructions.

3.2.1 The maximal twist

The configuration just discussed constitutes the last stage of orbifold twists at which we can "easily" follow the pattern of projections on all the three types of string construction. It represents also the maximal degree of Z_2 twisting corresponding to a supersymmetric configuration. As we will see, a further projection necessarily breaks supersymmetry. The vacuum appears supersymmetric only in certain dual phases, such as the perturbative heterotic representation. Non-perturbatively, supersymmetry is on the other hand broken. This means that, with further twisting, the theory is basically no more de-compactifiable: perturbative, i.e. decompactification, phases, represent only approximations in which part of the theory content and properties are lost, or hidden. This is what usually happens when one for instance pushes to infinity the size of a coordinate acted on by a Z_2 twist. The situation is the one of a "non-compact orbifold".

The further Z_2 twist we are going to consider is also the last that can be applied to this vacuum, which in this way attains its maximal degree of Z_2 twisting. This operation, and the configuration it leads to, appears rather differently, depending on the type of string approach. Let's see it first from the heterotic point of view. So far we are at the $\mathcal{N}_4 = 2$ level. The next step appears as a further reduction to four supercharges (corresponding to $\mathcal{N}_4 = 1$ supersymmetry). Of the previous projections, $Z_2^{(1)}$ and $Z_2^{(2)}$, only one was realized explicitly on the heterotic string, as a twist of four coordinates, say $\{x_5, x_6, x_7, x_8\}$. The further projection, $Z_2^{(3)}$, acts on another four coordinates, for instance $\{x_3, x_4, x_7, x_8\}$. In this way we generate a configuration in which the previous situation is replicated three times. When considered alone, the new projection would in fact behave like the previous one, and produce two non-perturbative sectors, with coupling parametrized by the moduli of a two-torus, in this case $\{x_5, x_6\}$: $T^{(5-6)}$, $U^{(5-6)}$. The product $Z_2^{(1)} \times Z_2^{(3)}$ leaves instead untwisted the torus $\{x_7, x_8\}$ and generates two non-perturbative sectors with couplings parametrized by the moduli $T^{(7-8)}$, $U^{(7-8)}$. Altogether, apart from the projection of states implied by the reduction of supersymmetry, the structure of the $\mathcal{N} = 2$ vacuum gets triplicated.

The symmetry of the action of the additional projection with respect to the previous ones suggests that the basic structure of the configuration, namely its repartition into three sectors, S, T, U, is preserved when passing to the less supersymmetric configuration. This phenomenon can be observed in the type II dual, that we discuss in detail in Appendix B. From the heterotic point of view, the states of these sectors come replicated ($\{T\} \to \{T^{(3-4)}, T^{(5-6)}, T^{(7-8)}\}$, $\{U\} \to \{U^{(3-4)}, U^{(5-6)}, U^{(7-8)}\}$). On the type II side we observe a triplication also of the "S" sector. However, as we discussed in Ref. [11], we are faced here to an artifact of the orbifold constructions, that by definition are built over a linearization of the string space into planes separated by the orbifold projections. The matter states are indeed charged under three sectors, S^i, T^j, U^k, but we can at most observe a double charge, as it appears on the type I dual side; from an analysis based on string-string duality, we learn that the states are in fact multi-charged for mutually non-perturbative sectors When one of the S^i, T^j, U^k sectors is at the weak coupling, the other two are at the strong coupling, and it doesn't make sense to ask what is this sort of "splitting" of the non-perturbative charge of the states: we simply observe that they have a perturbative index and one running on a strongly coupled part of the theory.

On the type I dual realization of this vacuum, besides a D9 branes sector we have now three D5 branes sectors and a replication of the non-perturbative sector into three sectors, whose couplings are parametrized by $U^{(3-4)}$, $U^{(5-6)}$, $U^{(7-8)}$.

A result of the combined action of these projections is that all the fields S^i, T^j and U^k are now twisted. This means that their vacuum expectation value is not anymore running, but fixed. We will see below, section 3.2.2, that minimization of entropy selects this value to be of order one, thereby implying also the identification of string and Planck scale (section 4). Nevertheless, for convenience here we continue with the generic notation S, T, U used so far, because it allows to better follow the functional structure of the configuration we are investigating. Twisting of the "coupling" moduli indeed suggests the non-decompactifiability of this vacuum. This, as discussed, would imply the breaking of supersymmetry. However, this property is not so directly evident: each dual construction is in fact by definition perturbatively constructed around a decompactification limit. The point is to see, with the help of string-string duality, whether this is a real decompactification, or just a singular, non-compact orbifold limit. An important argument in favour of this second situation is that, after the $Z_2^{(3)}$ projection is applied, the so-called "$\mathcal{N} = 2$ gauge beta-functions" are unavoidably non-vanishing. According to the analysis of Ref. [11], this means that there are hidden sectors at the strong coupling [24]. As a consequence, supersymmetry is actually non-perturbatively broken, due to gaugino condensation. Inspection of the type II string dual shows explicitly the instability of the $\mathcal{N}_4 = 1$ supersymmetric vacuum.

In order to construct the type II dual, it is not possible to proceed as with the heterotic and type I string, namely by keeping un-twisted some coordinates. On the type II side the "$\mathcal{N}_4 = 1$" vacuum looks rather differently: the new projection twists all the transverse coordinates, leaving no room for a "space-time". This however does not mean that a space-time does not exist: all non-twisted coordinates, therefore the space-time indices, are non-perturbative. Their volume is precisely related to the size of the coupling around which the

[24] We refer the reader to the cited work for a detailed discussion of this issue.

perturbative vacuum is expanded. After $Z_2^{(1)}$ and $Z_2^{(2)}$, the only possibility for applying a perturbative Z_2 twist is in fact to act on $\{x_1, x_2,\}$ and on two of the $\{x_3, \dots, x_8\}$ coordinates, already considered by the previous twists. These can be either the pair $\{x_3, x_4\}$ or $\{x_5, x_6\}$, or $\{x_7, x_8\}$. Which pair, is absolutely equivalent. We can chose $Z_2^{(3)}$ such that:

$$Z_2^{(3)} : \quad (x_1, x_2, x_3, x_4) \; \to \; (-x_1, -x_2, -x_3, -x_4) \,. \tag{3.32}$$

The other choices are anyway generated as $Z_2^{(3)} \times Z_2^{(2)}$ and $Z_2^{(3)} \times Z_2^{(2)} \times Z_2^{(1)}$. Assigning a twist to some coordinates is not enough in order to define an orbifold operation: the specification must be completed by an appropriate choice of "torsion coefficients". The analysis of this orbifold turns out to be easier at the fermionic point, where the world-sheet bosons of the conformal theory are realized through pairs of free fermions [50]. We leave to the appendix B a detailed discussion of the construction of this vacuum. There we see how the duality map with $\mathcal{N} = 2 \to \mathcal{N} \to 1$ heterotic theory imposes a choice of "GSO coefficients" that leads to the complete breaking of supersymmetry, and discuss, in appendix C, how the breaking is tuned by moduli which in this vacuum are frozen at the Planck scale. This is also therefore the scale of the supersymmetry breaking [25].

The reason why the breaking of space-time supersymmetry can be observed in a dual in which space-time is entirely non-perturbative relies on the unambiguous identification of the supersymmetry generators. More precisely, what on the type II side it is possible to see is the projection of the supersymmetry currents on the type II perturbative space. Target space supersymmetry is in fact realized in string theory through a set of currents whose representation is built out of the world-sheet degrees of freedom. For instance, in the case of free fermions in four dimensions, we have:

$$G(z) \; = \; \partial_z X^{\mu} \psi_{\mu} \, + \, \sum_i x^i y^i z^i \,, \tag{3.33}$$

and

$$G(\bar{z}) \; = \; \partial_{\bar{z}} X^{\mu} \bar{\psi}_{\mu} \, + \, \sum_i \bar{x}^i \bar{y}^i \bar{z}^i \,, \tag{3.34}$$

where the index i runs over the internal dimensions. At the $\mathcal{N}_4 = 2$ level it is possible to construct both the representations of the type II dual, namely the one in which space-time is perturbative, and the one in which space-time is non-perturbative., Tracing the representation of the supersymmetry currents in both these pictures allows us to identify them also when the $Z_2^{(3)}$ twist is applied. Although, strictly speaking, there is no simple one-to-one linear mapping between coordinates of dual constructions, the fact that the dual representations of the currents share a projection onto a subset of coordinates common to both, enables us to follow the fate of space-time supersymmetry anyway.

[25] Among the historical reasons for the search of low-energy supersymmetry are the related smallness of the cosmological constant and the stabilization in the renormalization of mass scales produced by supersymmetry. In our framework, the value of the cosmological constant will be justified in a completely different way (section 5), in the light of a different way of interpreting string amplitudes, discussed in section 4. Also the issue of stabilization of scales in this framework must be considered in a different way: masses are no more produced by a field-theory mechanism, and field theory is not the environment in which to investigate their running.

The analysis of the type II dual confirms that the matter states of this vacuum are indeed three replicas of the chiral fermions of the theory before the supersymmetry-breaking, $Z_2^{(3)}$ projection. In the type II construction their space-time spinor index runs non-perturbatively; they appear therefore as scalars. In total, we have three sets of bi-charged states in a $\mathbf{16}\times\mathbf{16}$. In the minimal, semi-freely acting configuration, they get reduced to three sets of $\mathbf{4}\times\mathbf{4}$ by the further Z_2 shifts, acting on the twisted planes. As it was the case of the $\mathcal{N}_4 = 2$ theory, on the type II side their charges are non-perturbative, and they misleadingly appear as $(\mathbf{16}, \mathbf{16}, \mathbf{16})$, reduced to $(\mathbf{4}, \mathbf{4}, \mathbf{4})$. The impression is that we have three families of three-charged states. However, this is only an artifact of the orbifold construction. From the heterotic point of view, namely, the vacuum in which gauge charges are visible, two sectors of each family are non-perturbative and, as previously mentioned, the structure of their contribution to threshold corrections is an indirect signal that they are at the strong coupling (see Ref. [11]). The situation is the following: *either* we explicitly see all the gauge sectors, on the type II side, but we don't see the gauge charges, *or* where we can explicitly construct currents and see gauge charges (the heterotic realization), we see the gauge sector, and the currents, corresponding to just one index born by the matter states: the other are non-perturbative and strongly coupled.

The type II realization appears to be a different "linearization", or linear representation, of the string space, in which the non-perturbative curvature has been "flattened" through an embedding in a higher number of (flat) coordinates, which goes together with a redundancy of states due to an artificial replication of some degrees of freedom. On the type II string, twisted states can only be represented as uncharged, free states. Their charges are in any case non-perturbative, and we cannot observe a "non-abelian gauge confinement". These gauge sectors appear as partially perturbative on the type I side. However, the type I vacuum, like the heterotic one, corresponds to an unstable phase of the theory: it appears as supersymmetric although it is not. Moreover, inspection of the gauge beta-functions reveals that they are positive. Therefore, although appearing as free states, the currents on the D-branes run to the strong coupling and the apparent gauge symmetries are broken by confinement.

Let's **summarize** the situation. The initial theory underwent three twists and now is essentially the following orbifold:

$$Z_2^{(1)} \times Z_2^{(2)} \times Z_2^{(3)} \, . \tag{3.35}$$

In terms of supercharges, the supersymmetry breaking pattern is:

$$32 \ \xrightarrow{Z_2^{(1)}} \ 16 \ \xrightarrow{Z_2^{(2)}} \ 8 \ \xrightarrow{Z_2^{(3)}} \ 0 \ \ (4 \text{ only perturbatively}) \, . \tag{3.36}$$

The "twisted sector" of the first projection gives rise to a non-trivial, rank 16 gauge group; the twisted sector of the second leads to the "creation" of one matter family, while after the third projection we have a replication by 3 of this family. The rank of each sector is then reduced by Z_2 shifts of the type discussed in Ref. [26, 16, 27], two per each complex plane. As a result, each $\mathbf{16}$ is reduced to $\mathbf{4}$. On the type II side one can explicitly see, besides the shifts, both the total breaking of supersymmetry and the doubling of sectors under which the

matter states are charged. The product of these operations leads precisely to the spreading into sectors that at the end of the day separate into weakly and strongly coupled, allowing us to interpret the matter states as quarks [26]. On the type I side, the states appear in an unstable phase, as free supersymmetric states of a confining gauge theory, while on the heterotic side they appear on the twisted sectors, and their gauge charges are partly non-perturbative, partly perturbative. The perturbative part is realized on the currents. Like the type I realization, also the heterotic vacuum appears to be an unstable phase, before flowing to confinement; they are indeed non-perturbatively singular, non-compact orbifolds. This reflects on the fact that, as also discussed in Ref. [11], both on the heterotic and type I side, perturbative and non-perturbative gauge sectors have opposite sign of the beta-function. This signals that, as the visible phase is confining, the hidden one is non-confining.

3.2.2 Origin of four dimensional space-time

The product (3.35) represents the maximal number of independent twists the theory can accommodate: a further twist would in fact superpose to the previous ones, and restore in some twisted sector the projected states. Therefore, further projections are allowed, but no further twists of coordinates. These twists allow us to distinguish between "space-time" and "internal" coordinates. While the first ones (the non-twisted) are free to expand, the twisted ones are "frozen". The reason is that the graviton, and as we will see the photon, live in the non-twisted coordinates. Precisely the fact that graviton and photon propagate along these coordinates, and therefore "stretch", expand the horizon, allows us to perceive these as our "space-time". We get therefore "a posteriori" the justification of our choice to analyze sectors and moduli from the point of view of a compactification to four dimensions.

The radius at which the internal coordinates are twisted is fixed by minimization of entropy to be of the order of the string scale. In order to understand this, let's consider an ordinary, bosonic lattice, that for simplicity we restrict to just one coordinate. The partition function consists of a sum of two terms: the un-twisted/projected and the twisted/projected part. Roughly, the un-twisted part reads:

$$\mathcal{Z} \sim \int \left\{ 1 + \sum_{m,n \neq 0,0} \mathrm{e}^{-\tau[(m/R)^2 + (nR)^2]} \right\} \times |\eta|^{-1} . \tag{3.37}$$

The η factor encodes the contribution of the oscillators that build up a tower of states over the ground momentum and winding. The term of interest for us, the one that varies, is the factor within brackets, $\{\}$. Even in the case a shift acts on this coordinate (as is in our case), what remains of the (broken) T-duality is enough to state that entropy possesses a minimum at around the string scale, $R_0 \sim \mathcal{O}(1)$. As the radius increases (resp. decreases beyond the extremal point), the probability of the states with non-zero winding (resp. momentum) number in fact decreases the more and more rapidly, while the probability of the momentum (resp. winding) states approaches the limit value:

$$P_m \sim \frac{\int \mathrm{e}^{-\tau \, m^2/R^2}}{\mathcal{Z}} \quad \stackrel{R \to \infty}{\longrightarrow} \quad \frac{\int 1}{\mathcal{Z}} . \tag{3.38}$$

[26] As we will discuss, the leptons show up as singlets inside quark multiplets.

Therefore, in the limit $R \to \infty$, for any fixed momentum number m_R, the tower of momentum states with momentum number $m < m_R$ "collapses" to nearly the same probability: many of the formerly well separated steps shrink to a "continuum" of states of nearly identical, maximal probability. Similarly goes for $R \to 0$ and the winding instead of momentum states. Therefore, as the radius departs from the extremal value $R = R_0$, either by increasing or by decreasing, the distribution of probabilities "collapses" toward two kinds of possibilities: half of the states tend to acquire the same, maximal probability, while the other half tend to disappear. From a well differentiated configuration, in which any energy level possessed a non-trivial probability, we move toward a situation of higher symmetry, and therefore higher entropy. These two decompactification limits lead either to a restoration of the broken symmetry, or to a configuration of purely projected, i.e. a non-freely acting orbifold, but without the additional states coming from the twisted sectors. For what concerns entropy, these two situations are equivalent.

3.2.3 In how many dimensions does non-perturbative String Theory live?

We have seen that, with the maximal twisting, supersymmetry is broken. The string space is therefore necessarily curved. For some respects, this may seem strange. As we also discuss in Appendix B and C, even with the maximal twist, not all the string coordinates are twisted. At most, we have twisted eight of them. Why should then not be possible to decompactify one of the non-twisted ones, and obtain anyway a flat space? The fact that the space gets curved at the maximal twisting, even though this does not involve all the coordinates, is a property of the orbifold constructions; they are a kind of singular spaces, with a geometry which is flat everywhere apart from some singular points. The curvature is somehow all "concentrated" on these points. Although from a global point of view orbifold spaces are curved, locally they are almost everywhere flat. By the way, this is the reason why one can have the impression of being able to decompactify them and build a consistent supergravity theory at the decompactification limit [27]. Supergravity is a part of field theory, it is related to differential geometry concepts, it is a local theory. The signal that something illegal has been done comes from the investigation of the *pure stringy*, non-perturbative properties of the vacuum under consideration. That with the $Z_2^{(1)} \times Z_2^{(2)} \times Z_2^{(3)}$ twisting the string orbifold space is curved, it can be clearly seen by considering the type II point of view. There, all the transverse coordinates are twisted. Not the coupling, which remains untwisted. Nevertheless, we are in the presence of a perturbative series in which, at any order in the expansion around the non-twisted coordinate(s), we have a fully twisted, curved space. This sums up therefore to a curved space; the perturbative contribution to the curvature cannot in fact be cancelled by a possible non-perturbative term: the terms should be cancelled order by order.

Besides the above mentioned twists/shifts, the only way left out to further minimize entropy is to apply further shifts along the non-twisted coordinates. How many are they? From the type II point of view, there are no further, un-twisted coordinates. But we know that they are there, "hidden" as longitudinal coordinates eaten in the light-cone gauge and in

[27] As remarked in Appendix C, this is also the case of Refs. [43, 44].

the coupling of the theory. Some of these coordinates appear on the heterotic/type I side as *two* transverse coordinates. If we count the total number of twisted coordinates by collecting the information coming from intersecting dual constructions, and the coordinates which are "hidden" in a certain construction and are explicitly realized in a dual construction, we get the impression that the underlying theory possesses 12 coordinates. For instance, on the heterotic side we have a four-dimensional space-time plus six internal, twisted coordinates, and a coupling. On the type II side we see eight twisted coordinates. We would therefore conclude that the two additional twisted coordinates correspond to the coupling of the heterotic dual. On the other hand, no supersymmetric 12-dimensional vacuum seems to exist, at least not in a flat space: the maximal dimension with these properties is 11. This seems therefore to be the number of dimensions in which non-perturbative string theory is natively defined.

Let's have a better look at the properties of supersymmetry. As is known, the supersymmetry algebra closes on the momentum operator. When applied to the vacuum, we have:

$$\left\{ Q, \bar{Q} \right\} \approx 2\, M \,. \tag{3.39}$$

The mass M can be viewed as an order parameter for the supersymmetry breaking. Alternatively, we can view its inverse as a length:

$$< \left\{ Q, \bar{Q} \right\} > \cong \frac{1}{R} \,. \tag{3.40}$$

For finite radius, supersymmetry is broken; it is restored at infinite radius. In string theory, the non-vanishing of the r.h.s. of equation 3.39 implies on the other hand the presence of a non-vanishing curvature of the vacuum, i.e. of space-time. Let's collect the informations so far obtained:

1. As soon as the string space is curved, supersymmetry is broken.
2. In the class of orbifolds, the phenomenon of curving the string space can only be partially and indirectly seen, through the comparison of dual constructions.
3. These constructions are built on a (perturbatively) flat, supersymmetric background: they provide therefore "linearizations" of the string space.
4. The maximal dimension of a supersymmetric theory on a flat background is 11.

All this suggests that, when supersymmetry is broken, we are in the presence of an eleven-dimensional *curved* background. Any, forcedly perturbative, explicit orbifold realization requires for its construction a linearization of the background. Since a 11-dimensional curved space can be embedded in a 12-dimensional flat space, we have the impression of an underlying 12-dimensional theory. However, this is only an artifact; in fact, we never see all these 12 flat coordinates at once: we infer their existence only by putting together all the pieces we can explicitly see. But this turns out to be misleading: the linearization is an artifact.

The 12 dimensional background is only fictitious, we need it only in order to describe the theory in terms of flat coordinates. At the string level, of these coordinates we see only a maximum of 10.

The counting of twisted and un-twisted coordinates has to be considered from this point of view. Trading the two "space-time" transverse coordinates on the type II side for the coupling coordinates of "M-theory", as we did in section 3.2.2, doesn't mean that we really have two such coordinates: the only thing we know is that, once linearized, the curved space looks 12 dimensional. Indeed, going to the type II picture is a trick enabling to explicitly see the instability of the $\mathcal{N}_4 = 1$ vacuum, by switching on the operation on the "hidden", non-perturbative part of the theory. As a matter of fact, we are however in the presence of a maximum of seven "twisted" coordinates, i.e. coordinates along which the degrees of freedom don't propagate, and four un-twisted ones, along which the degrees of freedom can propagate. By comparison of dual string vacua, we can see that there is room to accommodate two more "perturbative" Z_2 shifts: through the heterotic and/or type I realization, we can explicitly see only two transverse non-twisted coordinates, plus two longitudinal ones, along which no shift can act. It remains then one "internal", truly non-perturbative coordinate, to which no shift has yet been applied. This can only be indirectly investigated: if we try to explicitly realize it, it will appear as a set of two coordinates, giving the fake impression to have room for two independent shifts.

Let's now count the number of degrees of freedom of the matter states. We have three families, that for the moment are absolutely identical: each one contains $16 = 4 \times 4$ chiral fermions. These degrees of freedom are suitable to arrange in two doublets of two different $SU(2)$ subgroups of the symmetry group. Each doublet is therefore like the "up" and "down" of an $SU(2)$ doublet of the weak interactions, but this time with a multiplicity index 4. As we will see, this 4 will break into 3+1, the 3 corresponding to the three quark flavours, and the 1 (the singlet) to the lepton. The number of degrees of freedom is therefore the right one to fit into three families of leptons and quarks. However, so far all these fields are massless and charged under an $SU(2)$ symmetry. We will see how precisely shifts along the space-time coordinates lead to the breaking of parity of the weak interactions and to a non-vanishing mass for these particles.

3.2.4 The origin of masses for the matter states

Shifts applied to the "internal" coordinates reduce the symmetry group through mass liftings that, owing to the fact that the coordinates are also twisted, remain for ever fixed; in this specific case, at the Planck scale. Also the shifts acting on the space-time coordinates reduce further the rank of the symmetry group. However, the breaking in this case is obtained through mass shifts that depend on the length of the non-twisted coordinates, i.e. on the size of space-time. Therefore, the matter states "projected out" by these shifts are not thrown out from the spectrum of the low energy theory: they acquire a "weak" differentiation in their masses; the mass difference is inversely related to the scale of space-time.

We have seen that, before any shift in the space-time coordinates is applied, each twisted sector gives rise to four chiral matter fermions transforming in the **4** of a unitary gauge group.

The first Z_2-shift in the space-time breaks this symmetry, reducing the rank through a "level doubling" projection. On the heterotic string, this operation acts by further doubling the level of the gauge group realized on the currents [28]. The consequences of this operation are that: 1) half of the gauge group becomes massive; 2) half of the matter states become also massive. The initial symmetry is therefore broken to only one rank-2 unitary group, under which only half of matter transforms. The remaining matter degrees of freedom become massive. Since this phenomenon takes place at a scale related to the inverse of the space-time size, therefore lower than the Planck scale, field theory constitutes a good framework allowing to understand the fate of these degrees of freedom, leading to the creation of massive states. In field theory massive matter is made up of four degrees of freedom, corresponding to two chiral massless fields. In order to build up massive fields, the lifted matter degrees of freedom must combine with those that a priori were left massless by the shift. Namely, we have that the **4**, corresponding to $U(4)$, is broken by this shift to $\mathbf{2} + \not{\mathbf{2}}$:

$$U(4) \to U(2)_{(\mathrm{L})} \otimes \not{U}(\not{2})_{(\mathrm{R})} \,, \tag{3.41}$$

where the second symmetry factor is the broken one, with the corresponding bosons lifted to a non-vanishing mass. The two matter degrees of freedom charged under this group acquire a mass below the Planck scale, and combine with the two charged under $U(2)_{(\mathrm{L})}$. Therefore, of the initial fourfold degeneracy of massless matter degrees of freedom, we make up light massive matter, of which only the left-handed part feels an $U(2)$ symmetry, namely what survives of the initial symmetry. Indeed, the surviving group is the non-anomalous, traceless $SU(2)_{(\mathrm{L})}$ subgroup. As we will see, this group can be identified with the group of weak interactions. The chirality of weak interactions comes out therefore as a consequence of a shift in space-time. This had to be expected: the breaking of parity is in fact somehow like a free orbifold projection on the space-time. Together with the generation of non-vanishing particle masses and the breaking of parity, this shift also breaks the rotation symmetry of space-time, by separating the role of the two space-time transverse coordinates. We will come back to this issue in section 3.4.

It is legitimate to ask what is the mass scale of the gauge bosons of the "missing" $SU(2)$, namely whether there is a scale at which we should expect to observe an enhancement of symmetry. The answer is: there is no such a scale. The reason is that the scale of these bosons is simply T-dual, with respect to the Planck scale, to that of the masses of particles. Let's consider this shift as seen from the heterotic side. On the heterotic vacuum, matter states originate from the twisted sector, while the gauge bosons (the visible gauge group, the one involved in this operation) originate from the currents, in the untwisted sector of the theory. Similarly, on the type I side, gauge bosons and the charged states we are considering originate from D-branes sectors derived respectively from the untwisted, and the twisted orbifold sectors of the starting type II theory [29]. It is therefore clear that a shift on the string lattice lifts the masses of gauge bosons and those of matter states in a T-dual way.

[28] Notice that, once the four-dimensional space-time is included in the orbifold operations, owing to the rank reduction produced by the further shift made in this way possible, the gauge group realized on the currents becomes non-confining, inverting thereby the situation discussed at the end of section 3.2.1.

[29] The type II vacua are on the other hand not appropriate for the investigation of this phenomenon, because the gauge charges are non-perturbative. In any case, although in the form of just the Cartan

Since the scale of particle masses is below the Planck scale, the mass of these bosons is above the Planck scale; at such a scale, we are not anymore allowed to speak about "gauge bosons" or, in general, fields, in the way we normally intend them.

The shift just considered is the last "level-doubling, rank-reducing" projection allowed by this perturbative conformal approach. We are left however with two more coordinates which can accommodate a shift. One is a further coordinate of the extended space-time, the other is one of the twisted coordinates. From the heterotic point of view, this is an internal non-perturbative coordinate; just for simplicity, we can identify it with the 11-th coordinate of M-theory. A shift along the extended coordinate is somehow related to the breaking of the last perturbative symmetry we are left with, the $SU(2)_{(\mathrm{L})}$ symmetry. A shift along the 11-th coordinate breaks instead the underlying S-duality of the theory. This symmetry exchanges the role of strong and weak coupling.

With our approximation of Z_2 orbifolds it is however not possible to investigate these effects in a complete way: with the first shift, we have reached the boundary of the capabilities of the approximation we are using. As we said, Z_2 orbifolds constitute a good base on which to expand the string vacua of interest for us. But they are not able to account for the finest details: space appears factorized into orbifold planes, and we get the fake impression of a symmetry between these planes. As a consequence, the three matter families appear on the same footing. On the other hand, minimization of entropy tells us that this is not the minimum; the values at which the moduli of the theory are frozen lie not exactly at the orbifold point. They must be a bit "displaced", in a way that allows to distinguish and break the symmetry among orbifold planes. However, as we said, we expect the corrections to the orbifold approximation to be related to the scale of extended space-time. In order to investigate them, we will make use of global properties of the solution $\psi^{\min}$ of 2.18, as they can be explored with the help of thermodynamics.

3.3 Origin of the $SU(3)$ of QCD and low-energy spectrum

At the point we arrived, the low energy world appears made out of light fermionic matter (*no scalar* fields are present!), massless and massive gauge bosons. Matter states are charged with respect to bi-fundamental representations. More precisely, they transform in the $((\mathbf{2} \oplus \not{\mathbf{2}}), \mathbf{4})$, replicated in three families: $((\mathbf{2} \oplus \not{\mathbf{2}}), \mathbf{4}')$, $((\mathbf{2} \oplus \not{\mathbf{2}}), \mathbf{4}'')$, $((\mathbf{2} \oplus \not{\mathbf{2}}), \mathbf{4}''')$. As we discussed, a perturbative shift lifts the mass of the gauge bosons of the second $\mathbf{2}$ above the Planck scale and gives at the same time a light mass to the matter states. The $\mathbf{4}$ (i.e. $\mathbf{4}'$, $\mathbf{4}''$ and $\mathbf{4}'''$) are however at the strong coupling, and therefore broken symmetries. Indeed, outside of the orbifold point, we should better speak in terms of a larger symmetry, into which the matter degrees of freedom transform as $\mathbf{12} \supset \mathbf{4}' \oplus \mathbf{4}'' \oplus \mathbf{4}'''$, which is broken and at the strong coupling. The orbifold point constitutes a simplification, in which the $\mathbf{12}$ has been rigidly broken by the orbifold twists; the real configuration is a perturbation of this

---

subgroup of their symmetry group, gauge bosons and matter states arise from mirror constructions, related each other by the type II dual of the heterotic T-duality under consideration [11].

simplified situation, in which the bosons of the broken symmetry, which were acting among orbifold planes, acquire a non-vanishing mass, of the order of some power of the age of the Universe. On the other hand, minimization of entropy tells us that indeed the **12**, besides being broken into $\mathbf{4}' \oplus \mathbf{4}'' \oplus \mathbf{4}'''$, must be further broken to the minimal confining subgroup. Had we to represent this in terms of field theory, this would imply as unique possibility the breaking of **4** into $\mathbf{1} \oplus \mathbf{3}$, corresponding to $SU(4) \to U(1) \times SU(3)$. However, from the string point of view, strictly speaking the $SU(3)$ symmetry does not exist: the vacuum appears to be already at the strong coupling, and the only asymptotic states are $SU(3)$ singlets. More than talking about gauge symmetries, what we can do is to *count* the matter states/degrees of freedom, and *interpret* their multiplicities as due to symmetry transformation properties, as they would appear, in a field theory description, "from above", i.e. before flowing to the strong coupling. This is an artificial situation, that does not exist in the actual string realization. The string vacuum indeed describes the world as we observe it, namely, with quarks at the strong coupling. From: i) the counting of the matter degrees of freedom, ii) knowing that this sector is at the strong coupling as soon as we identify the "$(\mathbf{2} \oplus \not{\mathbf{2}})$" as the sector containing the group of weak interactions, and iii) the requirement of consistency with a field theory interpretation of the string states below the Planck scale [30], we derive that the only possibility is that this representation is broken to $U(1) \times SU(3)$. A further breaking would in fact lead to a negative beta-function, in contradiction with point ii). The operation that breaks the **4** into **3** and **1** corresponds to the "shift" along the 11-th coordinate we mentioned in section 3.2.4. This operation, that cannot be represented at the orbifold point, doesn't work as a "level doubling", but rather as a "Wilson line". Notice that, coming from the breaking of an $SU(4)$ symmetry, the $U(1)$ factor is traceless. This means that it acts by transforming with opposite phase states charged under $SU(3)$ and uncharged ones:

$$\begin{aligned} U(1)_\beta \, \varphi &= \mathrm{e}^{\mathrm{i}\beta} \, \varphi \,, \\ U(1)_\beta \, \varphi_a &= \mathrm{e}^{-\mathrm{i}\,\beta/3} \, \varphi_a \,, \qquad a \, \in \, \mathbf{3} \, \mathrm{of} \, SU(3) \,. \end{aligned} \tag{3.42}$$

At this point, we are left with one more extended space-time direction where to accommodate a shift, and therefore further break the symmetry.

Before considering the action of this last shift on the gauge groups, let's recall that, as we already mentioned and we will discuss more deeply in section 6, the real point of minimal entropy lies a bit displaced from the orbifold point. We expect the dominant configuration of the Universe to correspond to a weak perturbation of the orbifold vacuum, driven by moduli of the size of some power of the inverse of the age of the Universe. The mean value of the observables on the minimal entropy configuration should then be:

$$< \mathcal{O} >_{\psi^{\min}} \;\; \sim \;\; < \mathcal{O} >_{\psi^{\min}} \Big|_{Z_2} + \Delta < \mathcal{O} >_{\psi^{\min}} \,, \tag{3.43}$$

[30] For us, field theory is not something realized below the string scale. It is rather an approximation (that locally works), obtained by artificially considering the space-time coordinates as infinitely extended. As a consequence, the "field theory" scale is in our set up the scale of the compact, but non-twisted, string coordinates. In order to keep contact with the usual approach and technology, what we have to do is to send to infinity the size of space-time, and consider the <u>massless</u> spectrum of four dimensional string vacua.

with

$$\Delta < \mathcal{O} >_{\psi^{\min}} \; \approx \; \mathcal{O}\left(1/\mathcal{T}^p\right) \, , \qquad p > 0 \, . \tag{3.44}$$

Away from the orbifold point, the gauge groups are broken to $U(1)$'s. This is true for the groups which are at the weak coupling, namely those corresponding to the $(\mathbf{2} \oplus \not{\mathbf{2}})$. It doesn't hold for the $SU(3)$ "colour" symmetry, which is broken and at the strong coupling. Let's therefore concentrate on the $(\mathbf{2} \oplus \not{\mathbf{2}})$. This derives from an initial $U(4)$, then broken to $U(2) \otimes U(2)$ and "patched" to a single, massless $U(2)$, with the second $U(2)$ lifted at over-Planckian mass. Of this massless $U(2)$, acting on half of the chiral matter states, indeed only an $SU(2)$ subgroup survives, the traceless part. This can be identified with the group of the weak interactions, $SU(2)_{\text{w.i.}}$, which acts only on the left moving part of the matter states. A slight displacement away from the orbifold point, driven by moduli of the size of the inverse of the coordinates of the extended space-time ($\mathcal{O}(1/\mathcal{T}^p)$, $p < 1$), breaks this group to $U(1)$. This displacement does not break parity: it commutes with the operation that lifted the second $U(2)$ (indeed, one can think to perform this displacement first, and then the level-doubling shift on the $U(1) \times U(1)$ group).

Consider now the last shift operation, either at the $U(1)$ or at the extended $SU(2)_{\text{w.i.}}$ gauge symmetry point. This shift cannot act as a level-doubling projection, as it can be seen by looking from a heterotic point of view, in which the group from which the $SU(2)_{\text{w.i.}}$ of the weak interactions originates is realized on the currents. A further rank reduction through a level-doubling orbifold projection is forbidden by the embedding of the spin connection into the gauge group. The effect of this shift is that of lifting the mass of all the gauge bosons. This action cannot be followed in a heterotic construction, where the gauge fields arise from the currents, and cannot be "shifted". What happens can be observed on the type II side, where the gauge fields originate in the twisted sectors. From the type II point of view, it is clear that such a shift corresponds to making the orbifold projection to act freely. For a shift realized "on the momenta" of space-time, i.e. a "field theory shift", as the one we are considering, the mass of the bosons will be below the Planck mass:

$$m_{W^i} \; \overset{\approx}{\to} \; \frac{1}{2} \times \frac{1}{R} \, , \tag{3.45}$$

where R is the shifted coordinate. We will come back to discuss the mass of the $SU(2)_{\text{w.i.}}$ bosons after having gained some insight into the relation between perturbative mass expressions and the perturbatively resummed ones, in sections 6 and 7. Differently from the usual case, in which the matter states are lifted in a T-dual way to the bosons, here the mass of the matter states receives a shift of the same order as the gauge bosons. This phenomenon can be traced by looking at the type II dual realization. As discussed also in Ref. [11], on the type II side there are two mirror constructions: in one it is the (Cartan subgroup of the) gauge group which is explicitly realized on the twisted sectors; in the mirror we see instead the matter states. The two constructions are related by a T-duality in the internal coordinates, therefore not involving a transformation of the space-time coordinates in which this last shift acts. A shift that lifts gauge boson masses by acting on the momenta of one such coordinate remains a shift in the momenta also as viewed from the mirror construction, in which it is seen to lift the mass of matter states. What we expect is therefore that the mass

difference between the up and the down of a broken $SU(2)_{\rm w.i.}$ doublet is of the order of the broken boson masses. Namely, we expect the mass difference of the heaviest doublet, the one that sets the scale of the breaking of this symmetry in the matter sector, to correspond to the scale of the breaking of this symmetry in the gauge sector (m^i_W mass). We will come back for a more detailed discussion in sections 6 and 7. At the point of minimal entropy, the mass lift 3.45 is "combined" with a further mass shift produced by the breaking of the $SU(2)_{\rm w.i.}$ symmetry to $U(1)$. This is a "second order" deformation, driven by moduli as in 3.43, 3.44. The $U(1)$ boson will have therefore a mass in first approximation of the same order of the one of the W^+ and W^- bosons, plus a mass difference produced by a parity-preserving, and therefore left-right symmetric, displacement.

Leptons and quarks originate from the splitting of the $\mathbf{4}$ of $SU(4)$ (replicated in three families, $(\mathbf{4}', \mathbf{4}'', \mathbf{4}''')$), as $\mathbf{3} \oplus \mathbf{1}$ of $SU(3)$. All in all, we have therefore two $U(1)$ groups, or, better, if we also consider the Cartan of the lifted $SU(2)$ "Right", that we know to have acquired an over-Planckian mass, three $U(1)$s. Let's indicate their generators as $T_3^{\rm L}$, $T_3^{\rm R}$ and $\tilde{Y}$. The charge assignments along each of the three $(\mathbf{4} \otimes \mathbf{4}^i)$ are then:

$$\left(Q(T_3^{\rm L}) \oplus Q(T_3^{\rm R})\right) \otimes \left(Q(\tilde{Y})\right) \; = \; \left(\frac{1}{2}, -\frac{1}{2} \oplus \frac{1}{2}, -\frac{1}{2}\right) \otimes (\beta, -\beta/3, -\beta/3, -\beta/3) \, . \qquad (3.46)$$

$U(1)_{\tilde{Y}}$, the group 3.42, can be identified with some kind of hypercharge group. This is not exactly the hypercharge of the Standard Model: the charge assignments are here left-right symmetric, and $SU(2)$ invariant, because the two $SU(2)$, $SU(2)_{\rm (L)}$ and $SU(2)_{\rm (R)}$, of which the "Left" gives rise to the weak interactions group, are here factorized out. Since the hypercharge and the $SU(2)$ groups arise from mutually non-perturbative sectors, the hypercharge eigenstates are here singlets of the $SU(2)$, left *and* right, symmetries. This means that the phase refers to the sum of the up and down of each family:

$$\beta \;\propto\; \tilde{Y}_e + \tilde{Y}_\nu \, , \qquad (3.47)$$

$$\beta/3 \;\propto\; \tilde{Y}_{\rm up} + \tilde{Y}_{\rm down} \, , \qquad (3.48)$$

this for any family and colour. By using the standard normalization of Lie group generators, we set $|\beta| = 1$, and, by pure convention, $\beta = -1$. This group, the only surviving gauge group, whith a truly massless boson, can be identified with the electromagnetic group, and its charges with the "electric" charge assignment of the elementary particles of this vacuum. The problem is to see how the electric charge is distributed among the $SU(2)$ doublets, namely, according to which fraction the decompositions 3.47 and 3.48 are made. We know that, as a matter of fact, in nature this is not done "democratically": the electron has charge -1, and the neutrino is uncharged. The same charge difference applies also to colour triplets of quarks. This can be interpreted as the result of having taken a linear combination, more precisely the sum, of the $\tilde{Y}$, $T_3^{\rm L}$ and $T_3^{\rm R}$ charges. In our scenario, this distribution of the electric charge among the $SU(2)$ multiplets turns out to be required by entropy minimization. As we will discuss in detail in section 6.2.1, less interacting particles correspond to a less entropic configuration of the Universe. We can in fact view a particle as an object whose entropy is a way of "counting" the paths leading to its configuration. The number of paths is

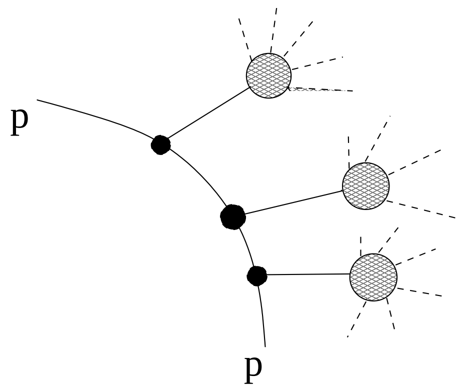


Figure 8: The more and stronger are the interactions of the particle p, the higher its entropy.

related to the strength of the interacting power of the particle: the more are the interactions the particle possesses, the more entropic is the physical configuration it corresponds to, because larger is the particle's "spread out" in the phase space. By "combining" the $U(1)$ charges we produce therefore a configuration of lower entropy, because we have minimized the interactions. This leads to the existence of a neutral particle (we will see that this must also be the lightest one), that we identify with the neutrino. The inclusion of the right-moving quantum numbers is not an "ad hoc" assignment, chosen in order to get the correct charge assignments: these are instead the consequence of the fact that, in this scenario, $\mathcal{T} \to 1$ is a limit of left-right symmetry restoration, where the dual, over-Planckian masses, come down to match the sub-Planckian ones. This is basically the meaning of having broken parity in a soft-way, by a shift along the extended coordinates. Therefore, no one of the charge assignments in 3.46 can explicitly break this symmetry [31]. Here is a point of difference with respect to the ordinary approach to the building of the electromagnetic group: in the usual approach, the hypercharge assignments explicitly break the $SU(2)$ and the parity symmetries.

Once the charge of the neutrino has been fixed, all other charges result correctly determined. Owing to the shift of the "hypercharge" with the $SU(2)$ charges, the electromagnetic current doesn't couple only with the "long range" force, the "photon", i.e. the massless field associated to the $U(1)_{\tilde{Y}}$ symmetry (indeed the sum of the up and down electric charges), but also with the massive neutral fields originating from the broken $SU(2)$s. The "Right" one is very massive, over the Planck scale, and therefore the corresponding interaction channel can be neglected: from a field theory point of view, "it does not exist" [32]. In section 7.6 we will discuss the masses of the broken $SU(2)$ and the "Left" $U(1)$ boson. The coupling of

[31]The case of the strong-weak coupling separation, related to the breaking of an S-duality, is different, because it involves a shift along coordinates which are also twisted, and therefore there is no restoration at the limit. The breaking is neither soft nor "spontaneous".

[32]This is a major point of difference with respect to the field theoretical "left-right" extensions of the Standard Model.

Figure 9: An effective neutrino electric charge is generated through higher order corrections.

the electromagnetic current will be derived and discussed in sections 6.2.5 and 6.2.9. The $SU(2)_{\text{w.i.}}$ coupling and the coupling of the neutral current with the under-Planckian massive neutral boson, the "Z" boson, will be discussed in sections 6.2.4 and 7.6 respectively.

It may seem disappointing that relations that we are used to consider associated to an exact symmetry, or concerning the exact vanishing of a charge, in this case the electric charge of the neutrino, appear in this scenario "softened", a consequence of minimization of entropy: there are no "conservation laws" strictly forbidding other solutions, and protecting these charges and masses. However, this is precisely what we should expect in a quantum scenario, and is accounted for in 2.31. In a way similar to what the well known path integral does, here *all* possible configurations contribute to build up the Universe as we observe it, although not all with equal weight: the most contributing ones are those of minimal entropy, followed by the ones which are close to minimal entropy and so on. Somehow the minimal entropy solution can be regarded as the "bare" configuration. The corrections to the minimal entropy configuration implied by 2.31 account for instance for the corrections to the electric charge of the neutrino: this vanishes in fact only as long as we consider the neutrino as a free, asymptotic state. Interactions generate indeed an effective non vanishing electric charge for this particle, as illustrated in figure 9. Expression 2.31 therefore contains the traditional field theory corrections to the bare states and parameters, and generalizes them to include also the cosmological configurations of space-time.

In section 6 we will discuss how the masses of matter states and of the $SU(2)_{\text{w.i.}}$ gauge bosons are computed in this scenario.

3.3.1 The fate of the Higgs field

If the origin of masses is a pure Stringy phenomenon, what about the Higgs field? The answer is simply that in String Theory there is no need of introducing such a field: its "raison d'être" was justified in a field theory context, in which one advocates the principle of "spontaneous breaking" of gauge symmetry in order to ensure renormalizability. But here, renormalizability of the theory holds just because it is string theory, which is not only renormalizable but finite. Masses are the consequence of the "microscopic" singularity of the string space. String theory provides therefore a unified description of space-time singularities: matter, i.e. particles and masses, are the result of twists and shifts on the string coordinates, they are generated by curving the string target space. In turn, they act

as “sources” of singularities for the “classical” geometry of space-time. On the other hand, the distinction between space-time in the traditional sense and the internal string space is only a matter of kind of singularity. Both the aspects find a common treatment within the entropy approach, which allows us to see the relation between the concentration or the spreading in the phase space of these singularities, to which we associate the concept of “state”, and a higher or lower probability for a certain configuration of the Universe to be realized [33].

In our framework of compact space-time, any field or particle has no more an expansion in terms of continuous Fourier modes: the momenta take value in a discrete lattice, and the description in terms of a continuum is an approximation the more and more appropriate as the volume of space-time becomes large. There is here no Higgs field, and the mass of a field or a particle corresponds to its lowest energy excitation. In this sense, it is a sort of “Casimir effect”. Indeed, in section 2.3 we anticipated that, in the case we want to translate the physical description in terms of a local, effective action, the masses appear to be “boundary effects”. The Higgs mechanism of field theory itself can here be considered “a way of effectively parametrizing the contribution of the boundary to the effective action” in the compact space-time. In such a sort of space, the gauge invariance is only locally preserved, and even this must be intended under the general condition that smoothness is only an asymptotic approximation. The Higgs mechanism, needed in field theory in order to cure the breaking of gauge invariance introduced by mass terms, is somehow the pull-back to the bulk, in terms of a density, i.e. a “field” depending on the point $\vec{x}$, of a term which, once integrated, should reproduce the global term produced by the existence of a boundary.

3.4 The breaking of Lorentz/rotation invariance

A common assumption about the fundamental laws of physics is that they are invariant under Lorentz transformations, and in particular under the subgroup of rotations. On the other hand, it is also common experience of everyday that, at least macroscopically, these symmetries are broken: by looking in different directions, we see different things. The same is true for time reversal and parity symmetries. For these two, however, it is known since long time that they are broken also at the microscopical level, namely in the very fundamental description of the physics of elementary particles. In the Standard Model this breaking is imposed “by hand”, while in some proposed extensions (left-right symmetric extensions) this comes out as the result of a spontaneous symmetry breaking, similar, and associated, to the one that gives rise to masses, via a Higgs mechanism. In our framework, masses

[33] Although not simply factorizable, the string space can anyway be viewed as a kind of fibered manifold with some coordinates extended up to the horizon, the “base”, and the other ones, the fiber, frozen at the Planck size, making up the “internal space”. Roughly speaking, the geometry of the internal space (the structure of its singularities, whose “statistics” is here selected by an entropy principle), determines the particle content of the theory, i.e. whether we have electrons and quarks, and photons etc..., in short what type of elementary modes propagate. The geometry of the external, extended space, the “base”, tells us instead about the number of these modes in the Universe: any such mode is a source for a singularity of space-time, and knowing the geometry of space-time is basically equivalent to knowing the statistics of these modes. This too is ruled by the same entropy principle.

are not generated as in field theory, and symmetries are not broken via Higgs mechanisms. Nevertheless, the functional 2.31 implies that mean values of observables are given by the superposition of several contributions, coming also from non-minimal-entropy configurations. The fact that their weight, as compared to the weight of the minimal ones, decreases with time provides in some way a kind of "spontaneous breaking" along the cooling down of the Universe, recovering therefore this idea in non-field theoretical terms (see discussion in section 4.1).

We have seen that in our framework the origin of masses is related to the breaking of parity and time-reversal, and that the same operation implies also the breaking of rotation invariance. Indeed, the entire analysis of the string orbifold vacua has been performed in the light-cone gauge, where the Lorentz symmetry is explicitly broken. However, the breaking due to a particular choice of gauge is not a real breaking: it is just a convenient representation, through a particular "slice", of an invariant construction. On the other hand, in our framework, what happens is that the Lorentz symmetry is really broken. The breaking is realized at two levels: 1) the breaking of the Lorentz boosts; 2) the breaking of the subgroup of space rotations. The Lorentz boosts are not a symmetry in a compact space-time, in which transformations in space and time correspond to an evolution of the Universe. Nevertheless, rotations in principle could remain a symmetry of space: they are not in contradiction with its compactness. However, here we discover that minimization of entropy not only implies the breaking of parities and a motion from smaller toward larger space-time volumes (and therefore the breaking of the special Lorentz transformations), but also the breaking of the symmetry of space under rotations. And this precisely at the same time as masses are produced.

This does not come unexpected: the very fact of placing a matter excitation somewhere in space breaks this invariance, selecting a preferred direction, in a way similar to the one a Higgs field breaks a gauge symmetry by selecting a position among the entire orbit of the symmetry group. In the average, owing to the presence of a lot of particles, existing as single sources of curvature or grouped into larger objects almost homogeneously distributed in the space, we can still say that the Universe is on a large scale invariant under rotation. Nevertheless, strictly speaking this is not a "pure" invariance. From a field theory point of view we would say that it is "spontaneously broken".

In our framework, the functional 2.31 is not limited to the "free" appearance of the fundamental laws of physics, but it contains the information about the actual configuration of the Universe, therefore the space-time *with* its energy and matter content, namely its "on-shell" configuration. We expect therefore that it accounts also for this "soft" breaking of the rotation invariance. It is therefore not a surprise that precisely the mechanism that in our framework substitutes the Higgs mechanism in giving rise to masses is also responsible for the breaking of the space rotation symmetry, through shifts acting on the transverse space coordinates: the one associated to the breaking of parity, and the one associed to the breaking of $SU(2)_{\rm w.i.}$. The fact that this breaking occurs in the matter sector is precisely related to the fact that it is only when states are massive that they can be localized, and therefore "do not occupy the entire space". Moreover, a breaking acting in the gravity sector would produce a non-vanishing mass for the graviton. Indeed, the only "breaking" experienced in

the gravity sector is the one regarding the Lorentz boosts, related to the compactness of space-time. This leads to the existence of a minimal energy for the graviton, as well as for the photon, which turn now out to be localized, but, as we discuss in the final remarks of section 7.6, within a region of space extended as much as the Universe itself. The amount of breaking of space rotations produced in our framework is of the same order of the particle masses: the symmetry is first broken when the lightest masses are generated, in agreement with the expectation that it is precisely the existence of localized ($\equiv$ massive) objects that in practice breaks the equivalence of any direction of observation, creating an anisotropy of the space. Anisotropies of the Universe are produced by different configurations of matter and radiation in various regions of the space: as we will see in the next sections, these phenomena are of order $\mathcal{O}\left(1/\mathcal{T}^p\right)$, $0 < p < 1/2$.

In our framework, which describes the physics "on shell", the microscopical and macroscopical descriptions of the world are therefore sewed together. This is related to the fact that we are deeply embedding statistics, and entropy, in the fundamental description. There is therefore no fracture between a fundamental, microscopical physics possessing certain symmetries, and a macroscopical world in which for some reasons these laws are violated: in our framework it turns out that time-reversal, space-parity, reversibility, symmetry under rotations are broken at a very fundamental level. The two levels of description are here unified in a seamless way.

3.5 The fate of the magnetic monopoles

Under the conditions of the scenario we are discussing, namely of a universe "enclosed" within a finite, compact space, also the issue of the existence of magnetic monopoles changes dramatically.

Magnetic monopoles can be of two kinds: the "classical" ones, namely those associated to a non-vanishing "bulk" magnetic charge that parallels the electric charge in a symmetric version of the Maxwell's equations, and the topological ones. In our scenario there are no "classical" monopoles: their existence would be possible only in the absence of an electromagnetic vector potential, what we have called the "photon" A_μ; their existence has therefore been ruled out as soon as we have discussed the existence and the masslessness of this field.

The first idea about the existence of magnetic monopoles in the classical sense (i.e. non-topological) originated by a request of symmetry: were not for the absence of magnetic charges, the Maxwell equations would be completely symmetric in the electric and magnetic field. However, the symmetry of these equations, preserved in empty space, is precisely spoiled by the presence of matter states that are also electrically charged. In our scenario, the description of the Universe is "on-shell" and the presence of matter comes out as "built-in": it cannot be disentangled from the existence of space itself. As the most often realized configuration of the universe is the one with the lowest amount of symmetry, it is not so surprising that, for the same reason for which charged matter states are generated, in correspondence with a local breaking of the invariance of space under $SO(3)$ rotations, also the symmetry of the Maxwell's equations is broken.

On the other hand, in this scenario there are no topological monopoles either. As all

vector fields are twisted (i.e. massive at the Planck scale or above it) with the only exception of the photon A_μ, propagating in the four-dimensional space time, and as this space-time dimensionality is electro-magnetically self-dual, the only possible topological monopoles would be those of the four-dimensional space coupled to the same photon field A_μ, namely, configurations à la t'Hooft and Polyakov or similar [34]. However, any such topological configuration is characterised by its being living in an infinitely-extended space: only in this way it is in fact possible to make compatible the existence of a p-form working as a "potential" $A_{(p)}$, defined as an analytic function in every point of the space, with the presence of a non-trivial magnetic flux. Therefore, here they cannot exist. As is well known, the magnetic flux through a surface can be computed as a loop integral of the vector potential. In the case of a surface enclosing a finite volume, the total flux is the sum of the loop integral circulated in both the opposite directions, so that it always trivially vanishes. However, things are different if the field has a non-trivial behaviour at infinity. At infinity we need just the circulation in one sense, because there is no "outside" from which field lines can "re-enter" in the space: if there is a non-vanishing circulation, there is a non-vanishing magnetic flux, and therefore also a non-vanishing magnetic charge. This however also means that, provided it exists, such a magnetic monopole is a highly non-localised object, with a magnetic field/vector potential such that the magnetic flux vanishes through any compact finite closed surface [35]. As a consequence, also the magnetic charge density point-wise vanishes at any location in the "bulk". Moreover, in our case we don't have a Higgs mechanism either. Furthermore, since the surface at infinity does not belong to any configuration of space-time, there is no smooth limit with a true restoration of the conditions at infinity allowing the existence of non-trivial topologies and homotopy groups. Light states with topological magnetic charges do not exist at all, not even approximately as the time becomes very large [36].

[34] for a review and references, see for instance [51, 52].

[35] Notice that the situation around the zero-dimensional point is equivalent to the one around the surface at infinity: if on one side the Dirac string can be considered as somehow the "dual" picture of the surface at infinity of the t'Hooft and Polyakov construction, in our scenario both infinity and the dimensionless point are excluded. Differential geometry and gauge theory are here only approximations.

[36] The situation is similar to the case of the volume of the group of translations and its identification with the regularized volume of space in the usual normalization of operators and amplitudes, completely absent in our scenario, something that leads to a different interpretation of string amplitudes as global quantities instead of densities, cfr. Section 4.

4 Effective theory and string corrections

Under the hypothesis of uniqueness of string theory, the various string constructions have to be considered as slices, at different points in the moduli space, of the same theory. Dual constructions corresponding to slices of overlapping regions of the moduli space. In order to investigate the properties of string theory in a certain region of its space. In the previous chapter we have discussed how, through the help of string duality, it is possible to analyze the structure of singularities in a class of orbifold constructions. We have also proposed a new formulation of the partition function, and discussed how this is more suited to the purpose of relating string computations with physical quantities and observables. The comparison of dual string constructions makes only sense once the expressions of the physical quantities are converted into common units, i.e. when, instead of being expressed in terms of the proper length of each string construction, they are expressed in terms of the Planck length or mass scale. Whenever one can write an effective action for the string light modes, this conversion corresponds to the passage from the string frame to the duality-invariant, so called Einstein's frame, also referred to as the "supergravity frame". In the case of compact space-time, the existence of an effective action is however not obvious. Traditionally, i.e. in an infinitely extended space time, the limit of infinite volume selects as elementary excitations of the propagating modes one of the two T-dual worlds, or towers of states whose quantum numbers correspond either to windings or to momenta. The other ones (therefore either the momenta or the windings) are "trivialized" to a constant contribution. In the case of compact space, at any volume both T-dual degrees of freedom are instead non-trivially present, and the selection of a configuration close to the infinite volume case is only possible when T-duality along the space-time coordinates is broken. The breaking of T-duality leads in fact to the choice of a "direction", or scale, for a field theory representation: either below or above the Planck scale, two situations no more equivalent.

As we have seen, in the class of configurations $\Psi^{\min}$, T-duality in the space-time coordinates is broken. This justifies the analysis in terms of orbifold dualities and comparison of dual constructions as we did in section 3.2, in terms of the "sectors" S, T and U of the theory, compared through the reduction of the effective actions to the common, Einstein's frame. At the $\mathcal{N}_4 = 2$ level, the last step at which we could still explicitly follow the pattern of dual constructions, for the heterotic string the relation between the string and the Planck scale is given by:

$$\mathrm{M}_{\mathrm{P}}^2 \equiv \mathrm{M}_{\mathrm{Het}}^2 \, \mathrm{Im}\, S^{-1} \,, \tag{4.1}$$

where $S = \chi + \mathrm{i}\mathrm{e}^{-2\phi}$, ϕ is the dilaton field.

On the type I side there are two such fields, called S and S', that parametrize the coupling of different D-branes sectors of the theory. Each sector has therefore its own "string length", to be rescaled to the Planck length. By inspecting string-string duality with the type II string, we have on the other hand learned that the vacuum of interest for us possesses at the $\mathcal{N}_4 = 2$ level three such "coupling fields", corresponding to three sectors of the theory (see also Ref. [11] for a more detailed discussion). This structure, namely the triplication of the string vacuum, is preserved when going to lower (super)symmetric, less entropic configurations, up to the minimal entropy one. We can therefore talk of three "waves", the

"S", "T" and "U" "wave" of the theory. Correspondingly, there are three "string scales" to be related to the Planck scale. In terms of the notation "S, T, U", these relations read:

$$\mathrm{M}_{\mathrm{P}}^2 \;=\; \mathrm{M}_{\mathrm{S}}^2 \, \mathrm{Im}\, S^{-1} \;\;=\;\; \mathrm{M}_{\mathrm{T}}^2 \, \mathrm{Im}\, T^{-1} \;\;=\;\; \mathrm{M}_{\mathrm{U}}^2 \, \mathrm{Im}\, U^{-1} \,. \tag{4.2}$$

As we have seen in section 3.2.2, the internal space is frozen at the self-dual radius, $\mathcal{O}(1)$ in string units. This implies that string and Planck scales are equivalent, as anticipated. From this equivalence, we conclude that the masses of the states lifted by the shifts acting along the internal coordinates are of the order of the Planck mass; these states are therefore black holes.

Owing to the breaking of T-duality along the space-time coordinates, realized in the class minimal entropy configurations ψ_V^{min}, it is possible to talk about "extended" space-time, along which degrees of freedom, representable as fields, can propagate. The solutions ψ_V^{min} are characterized by the existence of two massless modes, the graviton and the photon. They are free to move at the speed of light along the 4 non-twisted coordinates, the "space-time" of our Universe. We have seen that entropy increases with the volume of space-time, and that the system evolves toward configurations of increasing entropy. This agrees with the fact that the two massless modes propagate, thereby "extending" the volume of space-time. Moreover, as we have seen in section 3, what breaks parity along the space and time coordinates is precisely the mass lift produced by the shift that breaks T-duality. All these effects appear therefore to be tightly related, different aspects of the same phenomenon: owing to the consequent breaking of time reversal, the breaking of T-duality leads to the choice of an arrow for the time evolution. Indeed, only owing to the breaking of T-duality space-time volume expansion is not equivalent to space-time contraction.

Noteworthy is that not only T-duality (and time reversal) are broken, but are broken in a "soft" way. The "initial" configuration of the history of the Universe, at space-time volume $V = 1$, is in fact the one with the maximal amount of projections, that differentiate the spectrum at fixed volume. These projections are at each step of symmetry reduction "minimal", a condition required in order to maximize their number. T-duality turns therefore out to be broken by a coordinate shift, that does not project out states: it just shifts masses creating a gap between half of the states and the other half, their T-duals. Minimization of entropy imposes therefore not to make big operations of projections and go immediately to end, but to pass through the maximal differentiation and maximal amount of operations. The consequence is that the increase of entropy during evolution is the minimal possible out of the balance between increase and time-reversed decrease.

This ensures that, in the average, the system or, better, ψ^{min}, doesn't jump, it "passes through all the steps", and step by step the increase of entropy is the minimal one allowed. For instance, in a decay process the maximal probability is that of decaying to the "nearest" configuration, e.g. from the third family to second one and not directly to first one. Were it not so, were for instance T-duality "hardly" broken and all the corresponding gauge bosons lifted to infinite mass as in the case of a twist, the system would jump immediately, since the very beginning of its evolution, to its final configuration, and there would be no "history" in the Universe. Entropy would be immediately the maximal one, at any time $t = t_0 + \epsilon$ for every $\epsilon \geq \frac{1}{t_0}$, where $t_0 = \mathcal{O}(1)$ is the initial time.

Starting at the minimum of entropy at the minimum of space-time volume, space-time expands at the speed of light along the non-twisted coordinates. At any point of its evolution, it is bounded by a horizon defined by the hypersurface:

$$\mathcal{T}^2 \, = \, X_1^2 \, + \, X_2^2 \, + \, X_3^2 \, , \tag{4.3}$$

where $\mathcal{T}$ is the age of the Universe, and X_1, X_2, X_3 the space coordinates. Owing to the breaking of T-duality, we are allowed to talk about an effective action, in which the light modes of the theory, namely those whose mass is below the Planck scale, move in a space-time frame of coordinates larger than the Planck length. Heavier modes are integrated out and their existence manifests itself only through their contribution to the parameters of the effective theory. As explained in section 2.3, through an appropriate conversion of units we can identify the inverse of the age of the Universe with the "temperature" of the system, and label thereby the solutions of the class $\psi_V^{\rm min}$ with $\mathcal{T}$. At any point in the evolution, there is therefore an effective action for the light modes:

$$S_{\mathcal{T}} \, = \, \int_{[0,\mathcal{T}]} d^4 x \left(R + \Lambda + \ldots \right) \, , \tag{4.4}$$

where the integration region is a "ball" centered on the observer and bounded by the horizon. Expression 4.4 refers to the Einstein frame, where lengths are measured in Planck units. In the effective action, the integration is performed on a domain extended as much as the Universe up to the present-day horizon. The quantities appearing in the integrand are densities, and as such bear a dependence on the coordinates of space-time. From a physical point of view, points far away in space are far away also in time. As a matter of fact, however, one disentangles the space from the time dependence, so that the effective action of the local physics is obtained by extrapolating the physics valid at the point where the observer is located also to points located at a non-vanishing distance in space from the observer. One studies then the time dependence of these quantities. Although this does not correspond to the real way observations are made, it is convenient to take here too this point of view, because it is according to this that experimental data are "converted" into parameters of an effective action, the parameters to which we want to compare the predictions resulting from computations performed in our string scenario.

Expression 4.4 is what in our set up substitutes the traditional effective action. The metric of the background space has been considered flat: $\sqrt{-g} \, = \, 1$. The curvature of the space in this scenario has not to be considered an external input: we start with a flat metric in an empty space, and a curvature is generated by matter and the cosmological term. Corrections to the flat metric are therefore of second order, and are generated by the internal "dynamics" of the system.

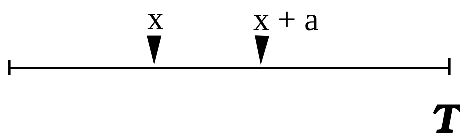
For what matters the local physics, there is no much difference between 4.4 and the usual effective action in an infinitely extended space-time, at least as long as the contribution of the boundary can be neglected. This is certainly true at our time of Universe "very large" as compared to the Planck scale. However, if we want to look at phenomena such as the cosmological constant, or the cosmological evolution of couplings and masses, the boundary enters heavily in the game. The boundary plays a role not only at the cosmological level:

in our scenario masses originate as "stringy boundary terms". In the action they enter as effective parameters, and must be considered as external inputs of a broken gauge theory.

As we discussed in Ref. [4], finiteness of the space-time volume implies a deep change of perspective, and leads to a slightly different interpretation of expression 4.4, as compared to the usual one. Indeed, here the effective action is by definition the action of *what we observe*: heavy modes are integrated out because they are black holes, and the domain of integration is finite, it possesses a well defined cut off. This explicitly breaks scale invariance, as well as invariance under space and time translations, and must be taken into account when comparing terms of the effective action, naturally related to the physical parameters we experimentally observe, with the results of computations performed on the string vacua.

Let's consider now a generic string computation, such as the renormalization of the coupling of a term of the effective action. This can be for instance an $F_{\mu\nu}F^{\mu\nu}$, R^2 term, or the renormalization of a constant, the "vacuum energy". On the string side, this is performed by inserting an appropriate operator in the partition function. By the latter we should mean the full partition function, which is in general not known. In the most favourable cases, it is only partially and perturbatively known. The accuracy in the approximation of computations therefore varies from case to case, and it depends also on the kind of term one wants to compute, according to whether it receives contributions from a finite or an infinite number of terms of a perturbative expansion. What however characterizes all these computations, is the special meaning they assume in our scenario, related to the way they must be compared with the effective action in the Einstein's frame: as opposite to the usual approach with an infinitely extended space-time, in our case they don't calculate anymore densities, but <u>global quantities</u>.

As we discussed in Ref. [4], the reason why in the traditional approach string computations produce densities, to be compared with the integrand appearing in the effective action, is that, in an infinitely extended space-time, there is a "gauge" freedom. It corresponds to the invariance under space-time translations. In any calculation there is therefore a redundancy, related to the fact that any quantity computed at the point "$\vec{x}$" is the same as at the point "$\vec{x}+\vec{a}$". In order to get rid of the "over-counting" due to this symmetry, one normalizes the computations by "fixing the gauge", i.e. dividing by the volume of the "orbit" of this symmetry $\equiv$ the volume of the space-time itself. Actually, since it is not possible to perform computations with a strictly infinite space-time, multiplying and dividing by infinity being a meaningless operation, the result is normally obtained through a procedure of "regularization" of the infinite: namely, one works with a space-time of volume V, supposed to be very big but anyway finite, and then takes the limit $V \to \infty$. In this kind of regularization, the volume of the space of translations is assumed to be V, and it is precisely the fact of dividing by V what at the end tells us that we have computed a density. In any such computation this normalization is implicitly assumed. In our case however, we do not assume the invariance under translations to be preserved for compact space-time, and indeed it is not: the situation we are describing is not the one of an ordinary "compactification": for us space-time is "absolute", is extended up to the "horizon". A translation of a point inside this space, $\vec{x} \to \vec{x} + \vec{a}$ is not a symmetry, being the boundary fixed. As illustrated below, the point $x + a$ lies closer than point x to the right boundary, corresponding to $\mathcal{T}$:



On the other hand, a translation of the boundary value of the target space string coordinate, $\mathcal{T}$, represents an evolution of the Universe, not a symmetry of the present-day effective theory: there is no "outside" space in which the coordinate $\mathcal{T}$ is embedded: at all the effects, there is no space-time beyond the horizon. In our framework, the volume of the group of translations is not V. Simply, this symmetry does not exist at all. There is therefore no over-counting, and what we compute is not a density, but a global value. In other words, what in the traditional interpretation is a space density, the value of a quantity at a certain point of space-time of the present-day Universe, in our case turns out to be a density in the "space" of the whole history of the Universe, the (global) value of this quantity at a certain point of its evolution [37].

As a consequence of the lack of invariance under translations, there is not only a change in the normalization of amplitudes, no more divided by the volume of space-time; the theory looses now also its invariance under reparametrization of the world-sheet coordinates. More precisely, the freedom in the reparametrization remains a basic property of the underlying string construction, however the comparison with the effective action must now take into account the actual size of the coordinates at which a certain result has been computed. The map between world-sheet and target-space coordinates is forcedly degenerate. In the traditional approach, owing to reparametrization invariance, we have the freedom to rescale coordinates in such a way to, roughly speaking, "identify" two target-space coordinates with the world-sheet ones. String amplitudes are then obtained by integrating over these coordinates up to their "horizon", corresponding to the string length. In practice, this corresponds to having shrunk a dimension-two subspace of space-time to the string proper size. In order to consistently compare the results with the terms of the effective action we must reproduce the same conditions on the coordinates. This is done by shrinking the space-time analogously to the world-sheet. Alternatively, we can account for this "asymmetry" of space-time by simply switching on an inverse Jacobian for the rescaling of coordinates on the string side, thereby multiplying the string result by a space-time "sub-volume" V_2. This volume is basically obtained by squaring the Jacobian corresponding to the boost of the time interval from the string proper length to the age of the Universe. If x_0 is the world sheet variable along the world sheet time coordinate, running in an interval of length $\ell_{\rm S}$, the target space time coordinate is $t \sim \mathcal{T} \times x_0 / \ell_{\rm S}$, where $\mathcal{T}$ is the age of the Universe. The Jacobian under consideration is $|\partial t / \partial x_0|$. In order to be compared with 4.4, any amplitude

[37] Notice that, in this interpretation of string coordinates, there is no "good" limit $V \to \infty$, if for "∞" we intend the ordinary situation in which there is invariance under translations. This symmetry appears in fact only strictly at the limit. The volume of the group of translations is some kind of "delta-function" supported at infinity. On the other hand, this is not a problem in our picture: infinite space-time does not really belong to the history of the universe, for which the horizon, although increasing, is always finite.

$\mathcal{A}$ computed on a string vacuum must therefore be rescaled by:

$$\mathcal{A} \to \left| \frac{\partial t}{\partial x_0} \right|^2 \mathcal{A} = \mathcal{T}^2 \mathcal{A} \,, \tag{4.5}$$

and, in order to be converted into a density, divided by a space-time volume factor, i.e. the missing volume of space-time translation group. The latter scales as $V_{\text{space-time}} \sim \mathcal{T}^4$. As a result,

$$\mathcal{A}(x) \sim \frac{\mathcal{A}}{\mathcal{T}^2} \,. \tag{4.6}$$

Therefore, *global* quantities, such as for instance entropy, are expected to have an overall scaling $\sim \mathcal{T}^2$, i.e. as the square of the age of the Universe, whereas *densities*, such as the mass/energy density of the Universe, or the cosmological constant, will have an overall scaling $\sim \mathcal{T}^{-2}$, i.e. as the inverse square of the age of the Universe.

4.1 A note on the asymptotic configuration

Having at hand the overall scaling of entropy in the class of the dominant configurations of the Universe, we can now justify the approximation of considering, for the practical purpose of describing the present-day physics, just the class of minimal entropy configurations. As we already pointed out, also "neighbouring" configurations, corresponding to a slightly modified lay-out of the degrees of freedom (particles and fields) within the extended space-time, give a non-negligible contribution to the functional 2.31, and therefore to the way the Universe appears, because they differ by a small increase of entropy. To this regard, it does not really matter that the weight in 2.31 is not so picked as a "delta-function" on the very minimal entropy configuration, because in the average the physics we experience is more or less the same.

Different is the case of entropy changes produced by a different structure of the "internal space" and in general of the operations which act also at the Planck scale level, such as the twists and shifts discussed in section 3. In this case, moving to more entropic configurations requires to switch on some "internal" moduli, and implies thereby an un-twisting of degrees of freedom. The volume of the phase space gets therefore increased. We have seen that in the minimal configuration entropy scales as in a black hole, $S^{\text{min}} = k\,\mathcal{T}^2$. $\mathcal{T}^2$ is the variable part, which is multiplied by the constant contribution of the internal space, k. When further degrees of freedom are un-frozen, the scaling as a function of the volume of space-time changes to an effectively higher power (the phase space acquires more directions). In general, any such increase of entropy can be parametrized by an effective exponent. We can therefore write entropy as:

$$S^{(p)} \approx k\,V^p \,, \qquad p \geq 2 \,, \tag{4.7}$$

where $p = 2$ for the minimal entropy configurations. The contribution to 2.31 of configurations with higher entropy $S^{(p)}$, as compared to configurations with lower entropy $S^{(q)}$, $q < p$, scales therefore as:

$$\frac{\mathrm{e}^{-V^p}}{\mathrm{e}^{-V^q}} \stackrel{V \to \infty}{\longrightarrow} 0 \,. \tag{4.8}$$

At large volumes (large age of the Universe) lower entropy configurations become therefore the more and more effectively representative of the real physical world. In some sense, as we mentioned in section 3.4, the progressive increasing of the dominancy, in the mean values of observables, of the lowest entropy configurations, in which there is the maximal amount of symmetry breaking, can be regarded as a sort of realization of a "spontaneous breaking of symmetry". Therefore, although in this class of configurations the symmetries are broken since the very beginning, in the average observables approach the more and more a broken-symmetry behaviour as time goes by and the Universe cools down.

5 The Universe as a Black Hole

The overall scaling of entropy as the surface of the space boundary of the Universe, exp. 4.5, suggests that the Universe itself can be viewed as a Black Hole [53, 54]. Indeed, the total energy at a certain time $\mathcal{T}$ of the history of the Universe is given by the integral of the energy density over the space volume of the Universe at time $\mathcal{T}$. From 4.6, we learn that this scales as:

$$\mathrm{E}(\mathcal{T}) \sim \int_{\mathcal{T}} d^3 \frac{1}{\mathcal{T}^2} \approx \mathcal{T} \, . \tag{5.1}$$

We will now discuss in detail the evaluation of the mass and energy content of the Universe, and confirm the conclusions of section 2.1, namely that the metric of the Universe is the one of a sphere; the boundary of space-time turns out to be at all the effects the boundary of a spherical black hole, our Universe.

5.1 The energy density and the cosmological constant

Let's start by computing the energy density of the Universe. When investigating the minimization of entropy, in section 3.2, we have seen that $\mathcal{N}_4 = 2$ was the last step at which the couplings of the theory appeared as moduli of the string space. There, the theory consists of three "sectors", corresponding to couplings parametrized by the moduli "S", "T" and "U". When the space is further "twisted", these moduli are frozen. Nevertheless, as we already pointed out in section 3.2.1, the less supersymmetric/entropic theory inherits the structure from $\mathcal{N}_4 = 2$. As we have seen, at the $\mathcal{N}_4 = 2$ level the heterotic realization, being constructed around a small/vanishing expectation value of $\mathrm{Im}\, S$, corresponds to a perturbative realization of the "S-sector". This is the "slice" of the theory which in natural way describes the coupling with gravity. We will call it the "S-picture". When going to the less entropic vacuum, by "S-picture" we have to intend a non-perturbative slice, whose perturbative limit would correspond to the heterotic construction.

By comparison of dual constructions at the extended supersymmetry level, it is possible to see the symmetry of the theory under exchange of the three sectors, symmetry which is then inherited by the $\mathcal{N}_4 = 0$ vacuum. In each sector, the energy of the Universe is given by the mean value of the identity operator; the total amount is given by the sum of the contributions of the three sectors. In the configuration of minimal entropy, the $S \leftrightarrow T \leftrightarrow U$ symmetry is softly broken by the action of the "shifts" responsible for the breaking of the strong-weak coupling S-duality, and for a full bunch of distinguished masses for the matter states (this issue will be considered more in detail in section 6). This breaking is tuned by the non-twisted coordinates, and therefore, owing to the breaking of T-duality, the corrections to mean values are expected to be of the order of (some power of) the inverse of the age of the Universe, as in 3.43, 3.44.

As a consequence, we expect that also the contributions to the energy of the Universe differ by some power of the inverse of its age:

$$\frac{E_i}{E_j} \approx 1 + \mathcal{O}\left(1/\mathcal{T}^{p_{ij}}\right) , \quad p_{ij} > 0 \, , \tag{5.2}$$

where E_i, E_j stay for E_S, E_T, E_U. These three contributions to the energy of the Universe differ in their interpretation. A closer look at the situation through comparison of the heterotic and type I pictures (see also Ref. [11]) reveals that the S, T and U pictures are related by S- and T-dualities, that exchange weakly and strongly coupled sectors (i.e. confining matter and non-confining particles, together with their gauge bosons, including the photon) and the "bulk", i.e. the gravity sector. Therefore, the mean value of the identity operator, when inserted in the gravity sector, computes the vacuum energy of the "empty space-time" in the "M-theory", or "String Theory" frame. In the other sectors, the mean value of the identity corresponds to the energy of matter (the confining part) and of "relativistic particles" respectively. The basic equivalence of these three contributions to the energy of the Universe is therefore a consequence of the symmetry between the "S-". "T-" and "U- waves" of the "dominant configuration"; symmetry which is eventually broken by the very minimization of entropy, but the breaking is "soft" as compared to the scale of the twisted space.

We want now to pass from the energies to the energy densities. As discussed in section 4, this passage introduces a dependence on the Jacobian of the transformation of two space-time coordinates from the string/Planck scale to the actual scale of space-time. Let's start by considering the contribution to the energy density in the "S-wave". This corresponds to the standard heterotic picture. According to 4.6, it is given by:

$$\begin{aligned} E_S|_{\text{Einstein frame}} &= \; < \mathbf{1} > |_{\text{Einstein frame}} = \\ &= \; < \mathbf{1} >_S |_{\text{string frame}} \times \frac{\kappa}{\mathcal{T}^2} \; = \; E_S|_{\text{string frame}} \times \frac{\kappa}{\mathcal{T}^2} \,, \end{aligned} \tag{5.3}$$

where $\kappa/\mathcal{T}^2$ is the Jacobian of the transformation from the string frame to the Einstein's frame, and κ a normalization constant. According to 5.2, $E_S|_{\text{string frame}}$ depends on time at the second order, and, owing to the fact that:

$$< \mathbf{1} >_S |_{\text{string frame}} + < \mathbf{1} >_T |_{\text{string frame}} + < \mathbf{1} >_S |_{\text{string frame}} \; = \; 3 \,, \tag{5.4}$$

we have also that:

$$E_S|_{\text{string frame}} \, + \, E_T|_{\text{string frame}} \, + \, E_S|_{\text{string frame}} \; = \; 3 \,. \tag{5.5}$$

The quantity 5.3 corresponds to the so called cosmological constant. In order to fix the normalization κ we must take into account that in passing from the string picture to the effective action term, besides the conversion from target space to world sheet of the two longitudinal coordinates, responsible for the two-volume factor, there is also a further space-time contraction due to the Z_2 orbifold projection. On the four (two transverse and two longitudinal) space-time coordinates act altogether two Z_2 shifts, which lead to a factor 4 in the conversion of a four volume. In our case, we have a two-volume, and the factor is 2. We obtain:

$$\Lambda(\mathcal{T}) \; \equiv \; E_S|_{\text{Einstein frame}} \, = \, \frac{2}{\mathcal{T}^2} \, \times \, \left[1 \, + \, \mathcal{O}\left(\frac{1}{\mathcal{T}^{q_S}} \right) \right] \,, \tag{5.6}$$

where the quantity within brackets accounts for the second order correction mentioned in 5.2, which can be positive or negative, i.e. contribute to slightly increase the value of the cosmological constant as compared to the other energies, or decrease it. In any case, the quantity 5.6 is here not a constant, but, through the age of the Universe, turns out to depend on time. For the sake of simplicity, we will anyway keep the usual terminology, and refer to it as to the "cosmological constant". According to 5.6, at present time the value of this parameter is:

$$\Lambda(\mathcal{T} = 10^{61}\mathrm{M}_{\mathrm{P}}^{-1}) \;\approx\; 10^{-122}\,\mathrm{M}_{\mathrm{P}}^{2}\,, \tag{5.7}$$

as it effectively seems to be suggested by the experimental observations [55, 56, 57].

Let's now introduce a "cosmological density", ρ_Λ, defined through:

$$8\pi\, G_{\mathrm{N}}\, \rho_\Lambda \;\equiv\; \Lambda \;=\; \frac{2}{\mathcal{T}^2} \,\times\, \left[1 \,+\, \mathcal{O}\left(\frac{1}{\mathcal{T}^{q_S}}\right)\right]\,. \tag{5.8}$$

In a similar way, we introduce two other densities, related to the energy density of the T- and U-"waves", ρ_m and ρ_r:

$$8\pi\, G_{\mathrm{N}}\, \rho_m \;\equiv\; E_T|_{\text{Einstein frame}} \;=\; \frac{2}{\mathcal{T}^2} \,\times\, \left[1 \,+\, \mathcal{O}\left(\frac{1}{\mathcal{T}^{q_T}}\right)\right]\,, \tag{5.9}$$

and

$$8\pi\, G_{\mathrm{N}}\, \rho_r \;\equiv\; E_U|_{\text{Einstein frame}} \;=\; \frac{2}{\mathcal{T}^2} \,\times\, \left[1 \,+\, \mathcal{O}\left(\frac{1}{\mathcal{T}^{q_U}}\right)\right]\,. \tag{5.10}$$

The correction terms in the brackets are of different signs, in order to sum up to a constant, in agreement with 5.5. Each one of the above equations can be written as:

$$\begin{aligned}
\rho_i \;&=\; 2 \times \frac{<\mathbf{1}>_i|_{\text{string frame}}}{3} \,\times\, \frac{R}{2G_{\mathrm{N}}\left[\frac{4}{3}\pi R^3\right]} \\
&=\; 2 \times \frac{<\mathbf{1}>_i|_{\text{string frame}}}{3} \,\times\, \frac{M_{\text{Schw.}}}{V}\,,
\end{aligned} \tag{5.11}$$

where $M_{\text{Schw.}} \equiv R/2G_{\mathrm{N}}$ is the "Schwarzschild mass", and V is the volume of the Universe up to the horizon, $R = \mathcal{T}$. If we sum up the three densities, taking into account 5.4 we obtain:

$$\rho \;\equiv\; \rho_\Lambda \,+\, \rho_m \,+\, \rho_r \;=\; \frac{2M_{\text{Schw.}}}{V}\,. \tag{5.12}$$

This agrees with the interpretation of the Universe as a spherical black hole: the mass/energy content corresponds to the Schwarzschild mass of a black hole of radius $R = \mathcal{T}$. To be more precise, the energy content is twice as much as the one of a black hole, because quantum space-time is an orbifold, and the volume is effectively reduced by a factor 2 with respect to the one of an ordinary smooth space.

The functional 2.31 tells us that also non-minimal entropy configurations contribute for some amount to the mean values. Although suppressed, and the more and more suppressed as time goes by, these contributions correct the relation 5.12, implying that the Universe is "almost" a Black Hole, but the exact relation between its radius and its total energy is modified by non-minimal vacua. In this sense, we can say, something not at all surprising, that the Universe is a "quantum Black Hole". The Schwarzschild relation, a classical relation, is violated by "quantum gravity" fluctuations. Indeed, if we consider the whole Universe as a fluctuation out of the vacuum, from the Heisenberg's Uncertainty Relations we get that its energy must satisfy a relation of the type:

$$\Delta M \geq \frac{1}{2\mathcal{T}}, \tag{5.13}$$

where the equality is saturated by the "classical", Schwarzschild mass, at $\mathcal{T} = 1$.

There is nothing to worry about considering ourselves at the border, or, depending on the point of view, at the center, of a black hole, our Universe: although we are intuitively led to consider black holes as small, extremely dense objects, indeed the mass of a black hole scales linearly with its radius, while the mass density scales with the (inverse) volume, i.e. the (inverse) cubic power of the radius. Therefore, the larger the black hole, the lower is its density. For a radius equal to the age of the Universe, matter inside is as rarefied as we see it to be in our Universe.

Owing to the interpretation of the Universe as a black hole, we can define a temperature of the Universe as the total energy divided by the entropy. The temperature turns out therefore to be proportional to the inverse of the age of the Universe:

$$T \stackrel{\text{def}}{=} k^{-1} \frac{\mathrm{E}(\mathcal{T})}{\mathcal{S}(\mathcal{T})} = (4\pi k)^{-1} \frac{1}{\mathcal{T}}, \tag{5.14}$$

where $k \approx 8,62 \times 10^{-5} \mathrm{eV} K^{-1}$ is the Boltzmann constant, and the normalization has been fixed according to the usual black holes thermodynamics [53, 54], relation 2.24. The present value of the age of the Universe, converted in mass units, is $\mathcal{T} \sim 10^{61} \mathrm{M}_{\mathrm{P}}^{-1}$. At present time, the temperature of the Universe is therefore:

$$T \approx 1,1 \times 10^{-29}\, {}^{0}K . \tag{5.15}$$

This temperature, for any practical purpose indistinguishable from the absolute zero [38], has not to be confused with that of the CMB radiation ($\sim 3^{0}K$), that we will discuss in detail in section 9.1.

Not much surprisingly, the scaling of the temperature as the inverse of the age of the Universe is the same as the one found in standard cosmology, at ages sufficiently away from the Big Bang. At large times (large as compared to the Planck scale) the mean configuration of the Universe arising in our scenario can in fact be approximated by classical geometry

[38] The temperature was around one Kelvin at a time in Planck units 10^{29} times earlier than today, i.e. at $\mathcal{T} \sim 10^{33} \mathrm{M}_{\mathrm{P}}^{-1}$. Since 1 yr $\sim 10^{51} \mathrm{M}_{\mathrm{P}}^{-1}$, this temperature corresponds to the age $\mathcal{T}_0 \sim 10^{-19}$yr $\sim (3\times)10^{-12}$sec.

and a standard cosmological scenario. The difference is that for us this is a "mean value" of a deeply quantum scenario, where deviations from the dominant solution are weighted by something completely external to a field theory description. The scaling behaviour of energy/mass densities and of the temperature is therefore not at all a trivial result. In section 2.3 we proposed the functional 2.18 as the "generating function" of the theory, and we discussed how this "contains" also the ordinary path integral, providing therefore its generalization to quantum gravity. At the base of the connection to standard quantum mechanics and the ordinary path integral there were the identification of the temperature with the inverse of the time scale, and the property that in our framework the action of quantum gravity doesn't contain a potential term: it consists only of kinetic terms. We discussed how mass terms should arise as boundary terms, as a consequence of the finite extension of space-time, and the consequent non-vanishing of the variation at the boundary. This was a "semiclassical" way of thinking. Now that we have the tools to investigate the properties of the configuration, $\psi^{\min}$, which dominates in 2.18, we can check that, indeed, the mass/energy densities behave, as predicted by 2.18, as boundary terms.

5.2 The solution of the FRW equations

In section 2 we investigated the geometry of the Universe by following the path of light rays expanding at the speed of light. We saw how the progressive "curving" of the space is produced as a consequence of the finiteness of the speed of light, and discussed the implications on our way of measuring and making observations. We saw how this results in the geometry of a sphere. In the previous paragraph we have then seen that the energy density and the scaling of entropy, as implied by the dominant configuration, agree with the interpretation of the Universe as a black hole. It seems therefore that, as the Universe evolves, its mean configuration $\psi^{\min}$ tends to a "classical" description: the energy density and the curvature of space-time decrease toward a flat limit.

Here we want to discuss this issue from the point of view of a cosmological solution of the Einstein's equations. Let's consider the class $\{\psi^{\min}(\mathcal{T})\}$ of solutions which differ from the absolute minimal entropy configuration at time $\mathcal{T}$ just in the way matter and fields occupy space-time. For what we discussed in section 4.1, this difference is very mild as compared to differentiations in the twisted space, which instead lead to a different spectrum, i.e. field and matter content of the theory. In all these configurations entropy is no much higher than that of $\psi^{\min}(\mathcal{T})$. They contribute to the sum 2.18 with almost the same weight, and, to the purpose of establishing a contact with an average, classical geometry, we can think at them as at fluctuations centered on a "mean" configuration $< \psi^{\min}(\mathcal{T}) >$. For what we said, it is reasonable to suppose that this configuration admits a description in terms of Robertson-Walker metric, i.e. a classical metric of the type:

$$ds^2 = dt^2 - R^2(t) \left[\frac{dr^2}{1 - kr^2} + r^2(d\theta^2 + \sin^2\theta \, d\phi^2) \right] , \tag{5.16}$$

where for us $t \equiv \mathcal{T}$, and $r \leq 1$. For what we said, the metric should correspond to a closed Universe, $k = 1$. Under the assumption of perfect fluid for the energy-momentum tensor,

the Einstein's equations lead to:

$$\left(\frac{\dot{R}}{R} \right)^2 \; = \, - \, \frac{k}{R^2} \, + \, \left\{ \frac{8\pi \, G_{\rm N} \, \rho}{3} \, + \, \frac{\Lambda}{3} \right\} \, , \tag{5.17}$$

where we have collected within brackets the contribution of the stress-energy tensor and of the cosmological term. According to our previous discussion, the stress-energy contribution consists of two terms ($\rho = \rho_m + \rho_r$), each one of the same order of the Λ-term. Inserting the values given in 5.8, 5.9, 5.10 with the "Ansatz" $R \, = \, \mathcal{T}$, and summing up, we obtain:

$$\left(\frac{\dot{R}}{R} \right)^2 \; = \, - \, \frac{(k=1)}{R^2} \, + \, \left\{ \frac{2}{R^2} \right\} \;\; = \, \frac{1}{R^2} \, . \tag{5.18}$$

The equation is solved by $R \, = \, t$, consistently with our Ansatz. This confirms that the dominant configuration corresponds to a spherical Robertson-Walker metric, describing a Universe bounded by a horizon expanding at the speed of light. If instead of the densities we introduce as usual the quantities Ω_m, Ω_r, Ω_v, i.e. the densities rescaled in units of the Hubble parameter $H \, \equiv \, (\dot{R}/R)$, the Hubble equation 5.17 becomes trivially:

$$0 \, = \, \frac{k \, (=1)}{R^2} \, = \, H^2 \, (\Omega_m \, + \, \Omega_r \, + \, \Omega_v \, - \, 1) \; , \tag{5.19}$$

where we dropped the label "$_0$", which usually indicates the current, present-day value, from the Hubble parameter: this equation is now valid at any time, and is trivially solved by $\Omega_m \, + \, \Omega_r \, + \, \Omega_v \, = \, 2$ and $\dot{R} \, = \, 1$.

Besides the Hubble equation, it is common to derive further equations of motion, by imposing energy-momentum conservation to the Roberts-Walker solution of the Einstein's equations. In the present case however no further equation can be derived: energy-momentum conservation remains valid as a "local" law. At the cosmological scale, energy is not conserved: $E_{tot} \, \propto \, R \, = \, \mathcal{T}$.

⋆⋆⋆

The comparison of our results with experimental data, as we did in eq. 5.7, contains a possible weak point. Experimental data are given as a result of a process of interpretation of certain measurements, for instance through a series of interpolations of parameters. All this is consistently done within a well defined theoretical framework. Usually, one takes a "conservative" attitude and lets the interpolations to run in a class of models. However, this is always done within a finite class of models. In principle, we are not allowed to compare theoretical predictions with numbers obtained through the elaboration of measurements in a different theoretical framework. In general, this doesn't make any sense.

However, in the present case this comparison is not meaningless, and this not on the base of theoretical grounds: the reason is that, for what concerns the time dependence of cosmic parameters and energy densities, the solution we are proposing does not behave, at present time, much differently from the "classical" cosmological models usually considered in the theoretical extrapolations from the experimental measurements. The rate of variation

of energy density is in fact: $\dot{\rho} \sim \partial(1/R^2)/\partial\mathcal{T} = 1/\mathcal{T}^3 = 1/R^3$. The values of the three kinds of densities can therefore be approximated by a constant within a wide range of time. For instance, as long as the accuracy of measurements does not go beyond the order of magnitude, these densities can be assumed to be constant within a range of several billions of years. For the purpose of testing the statements and conclusions of the present analysis, the use of the known experimental data about the cosmological constant, derived within the framework of a Robertson-Walker Universe with constant densities, is therefore justified.

⋆⋆⋆

A Universe evolving according to eq. 5.18 is not accelerated: $\dot{R} = 1$ and $\ddot{R} = 0$. Owing to the existence of an effective Robertson-Walker description, the red-shift can be computed as usual. We have:

$$1 + z = \frac{\nu_1}{\nu_2} = \frac{R_2}{R_1} = \frac{\mathcal{T}_2}{\mathcal{T}_1}, \tag{5.20}$$

where ν_1 is the frequency of the emitted light, ν_2 the frequency which is observed, and R_1, R_2 are respectively the scale factor for the emitter and the observer. $R = \mathcal{T}$ is precisely the statement that the expansion is not accelerated. Expression 5.20 however accounts for just the "bare" red-shift, namely the part due to the expansion of the Universe: it does not account for further corrections coming from the time dependence of masses. Usually, this effect is not taken into account, because in the standard scenarios masses are assumed to be constant. As we will discuss, in our scenario they depend instead on the age of the Universe. A change in the values of masses reflects in a change of the atomic energy levels, and therefore in a change of the emitted frequencies. In order to properly include this effect in the computation of the red-shift, we must therefore discuss the time dependence of masses. This was partly done in Ref. [2], where we obtained in a "heuristic" way an approximated behaviour for the mass of stable matter, the one that indeed defines the "center of mass scale". We are now in the position to better understand this phenomenon, that we will completely rediscuss in the next sections. We will see that, once the observed frequencies in expression 5.20 are corrected to include also the change in the scale of energies, the scaling of the emitted to observed frequency ratio is not anymore proportional to the ratio of the corresponding ages of the Universe. Since the conclusions about the rate of expansion are precisely derived by comparing red-shift data of objects located at a certain space-time distance from each other, this explains why the expansion *appears to be accelerated.*

5.3 Back to the partition function

For what matters any computation in a non-supersymmetric vacuum, the expression 3.8 we have proposed for the partition function works definitely in the way a partition function should. However, it leaves unsolved the problem of correctly dealing with supersymmetric vacua. Owing to the properties of supersymmetry algebra, 3.39, supersymmetry provides a mechanism of cancellation of the vacuum energy, and this must in some way reflect in the partition function. It is like to say that a supersymmetric vacuum is geometrically flat:

supersymmetry cancels the ground curvature of space-time. That's why, inspired by 3.39, the partition function so far considered in the string literature has been the one based on a sum weighted with the eigenvalues of the operator $(-1)^F$, where F is the supersymmetry charge sign operator. In this way, supersymmetric partners contribute with opposite sign to the partition function, and the vacuum energy, obtained as the mean value of the constant operator inserted in the partition function, vanishes for unbroken supersymmetry, reproducing what we expect from 3.39. On the other hand, we have seen that this expression fails to give a correct evaluation of entropy. There seems therefore to be a contradiction in any attempt to define a correct partition function is a superstring scenario. Our aim is to discuss here a possible solution to this puzzle.

Through expression 3.39, the supersymmetry algebra indeed suggests that the mass on the right hand side, in all an order respects a parameter for the supersymmetry breaking, could be interpreted as the inverse of the length of a coordinate of the theory. This coordinate refers to an internal, extra dimension, or, perhaps more appropriately, a curvature, i.e. a function collecting the contribution of several coordinates, perturbative as well as non-perturbative. If we call $\mathcal{Z}_1$ and $\mathcal{Z}_2$ respectively the two contributions, corresponding to the eigenstates of the supersymmetry charge operator -1^F with eigenvalues ± 1, the traditional string partition function reads:

$$\mathcal{Z} \equiv \mathcal{Z}_- = \mathcal{Z}_1 - \mathcal{Z}_2 \,. \tag{5.21}$$

What we have proposed, for practical computations in the case of broken supersymmetry, is instead:

$$\mathcal{Z}_+ = \mathcal{Z}_1 + \mathcal{Z}_2 \,. \tag{5.22}$$

As we discussed, the environment in which string corrections are computed is in any case the one of broken supersymmetry. Once computed through either 5.21 or 5.22, amplitudes must be converted to the duality invariant "Einstein's frame", in which the strength of the string coupling is reabsorbed into a redefinition of the mass/length scale. In general, the partition function is a perturbative piece of an intrinsically wider theory, computed around a limit value of the string coupling, and string corrections consist of several contributions, originating from different sectors of the theory (bulk, branes...). The normalization of one of these pieces, when considered alone, is irrelevant, because it is just a linearized piece of an expansion around a "coupling" whose full, non-perturbative expression, is not known. From a higher dimensional point of view, the latter is on the other hand itself a coordinate.

As we discussed in section 3.2.3, it seems that we are in the presence of a theory living in a 11-dimensional space. This space is either flat (supersymmetric case) or curved. In this second case, it can be linearly represented only through a set of sub-dimensional embeddings in a 12-dimensional flat space. Our suggestion is that $\mathcal{Z}_-$, which realizes an implementation of 3.39, and therefore of 3.40, is somehow a parametrization of the curvature of the space. If we switch on the neglected normalization of the partition function, we must multiply expression 5.22 by the coupling of the theory, whose expression involves also the "coordinate" $\mathcal{Z}_-$. The full partition function should therefore be proportional to:

$$\mathcal{Z} \cong \mathcal{Z}_+ \times \mathcal{Z}_- \,. \tag{5.23}$$

When supersymmetry is broken, as is always the case when we compute threshold corrections of effective terms except from the vacuum energy, we always go to a region of the moduli space of the theory in which $\mathcal{Z}_-$ is frozen to a certain finite value. This results in a change of the overall normalization of the partition function. The latter is then implicitly reabsorbed in a redefinition of the coupling, and therefore "hidden" in the transformation from the string frame to the Einstein's frame. It can therefore be neglected. The traditional approach corresponds on the other hand to reabsorbing in the overall normalization not the piece $\mathcal{Z}_-$ but $\mathcal{Z}_+$. Again, for the practical purpose of computing threshold corrections, the two approaches lead in practice to the same result.

Precisely the fact that, in the breaking of $\mathcal{N}_4 = 2$ supersymmetry to $\mathcal{N}_4 = 1$, the dilaton and the other "coupling" fields get twisted, is a signal that a non-vanishing curvature of the string space has been generated. As we discussed in section 3.2.1, this means that, even in the case of infinite volume, we are in a situation of non-compact orbifold. A non-running value for the vacuum expectation value of these fields, i.e. a non trivial, fixed value of the corresponding couplings, is the necessary condition for the normalization of the partition function and the vacuum energy, and implies the breaking of supersymmetry:

$$\begin{aligned} &\text{broken susy} \quad \Leftrightarrow \quad \mathcal{Z}_- \neq 0 \\ &\quad \Leftrightarrow \quad < S >=< S >_0 \,, < T >=< T >_0 \,, < U >=< U >_0 \neq 0 \;\, [\sim \mathcal{O}(1)\,] \,. \end{aligned} \tag{5.24}$$

Simply assigning a non-vanishing value to the coupling does not imply the generation of a non-vanishing curvature of the string space: the corresponding field must also be twisted.

In the orbifold language, this is implemented by the fact that, whenever the coupling field is "explicitated" by going to a dual construction, the corresponding perturbative geometric field appears as a volume of a two-dimensional space. This phenomenon can be observed for reduced supersymmetry (for maximal supersymmetry, there is just the type II string construction). Consider for instance the eleventh coordinate of M-theory, that should correspond to the dilaton of the heterotic string. In the type II orbifold constructions (K3 orbifold compactifications), the heterotic coupling corresponds to a two-torus volume. Considering that this two-dimensional space corresponds, from the heterotic point of view, to "extra-coordinates", one would say that, in order to realize all these degrees of freedom, the full underlying theory should be (at least) twelve-dimensional. However, as we already discussed in section 3.2.3, this is only an artifact of the linearization implied by the orbifold construction, and it means that the simple compactification on a circle is not enough, we need also an additional "curvature coordinate" in order to parametrize a truly curved space.

From the type II dual we learn that supersymmetry is not restored by a simple decompactification: the string space is twisted [39]. Flatness of the string space is broken by a

[39]In some type II/heterotic duality identifications, the heterotic coupling is said to correspond to un-twisted coordinates of the type II string. This however does not change the terms of the problem: in the artifacts of the flattening implied by the orbifold constructions, part of the curvature may be "displaced", or referred, to some or some other coordinates. This "rigid" distribution of the twists, basically dictated by the need of recovering a description in terms of supergravity fields referring to the same space-time dimensionality for both the dual constructions, may induce to misleadingly conclusions. The intrinsic twisted nature of the

"twist" of coordinates that fixes them to the Planck scale. As a consequence, the supersymmetric partners of the low-energy states are boosted to black holes (mass at/above the Planck scale). In a situation of supersymmetry restoration, they should come down to the same mass as the visible world, and space should become "flat". However, this is only possible when the twist is "unfrozen" and we can take a decompactification limit, such as for instance the M-theory limit. Otherwise, at the decompactification limit the space becomes only locally flat (non-compact orbifold). In order to get a true flat space we must take out the "point at infinity" (the "twist"), which closes the geometry to a curved space.

The large volume limit of our Universe is precisely like a non-compact orbifold: at infinity the space looks flat, but the total energy is extremely big (it goes, as in a black hole, with the radius, therefore to infinity), and this because we do not "un-twisted" the space-time, as we have to do in string theory, in order to get the M-theory limit.

5.4 A note on matter-light-gravity duality

In this work, we have started our analysis by considering the effects of the finiteness of the speed of light on the geometric properties of space and time in the Universe. Without imposing any further condition besides bounding the Universe to a finitely extended causal region, we arrived to an entire quantum gravity theory, which predicts the existence of all three forms of energy. Indeed, the curvature of space-time we derived in section 2.1 by considering the paths of light rays and the horizon/origin equivalence, turns out to account for the *full* curvature, the sum of the three terms 5.8, 5.9, 5.10. It therefore already contains the information that the Universe must possess all the three forms of energy. It seems that the curvature originating from finiteness of the speed of light indeed "generates" the Universe as we see it, including matter. How is this possible? Namely, how can happen that a space-time, for the simple fact of existing, "creates" light and matter?

General Relativity tells us how a gravitational field is generated by energy, not only as matter but in any of its forms. This tells us that a photon, regarded as a wave packet carrying energy, not only is affected by a gravitational field, but it also generates a gravitational field. Since the time of relativistic quantum mechanics we also know that colliding photons can generate matter-antimatter pairs. Through CP asymmetric decays, matter ends up to prevail over antimatter. Putting together what General Relativity and Relativistic Quantum Mechanics tell us, we can therefore understand how a Universe fulfilled with photons ends up to possess all the forms of energy: radiation, matter, and gravity.

Here, we have seen that a space-time of finite size is necessarily curved, in the sense that it appears to us to be curved. Here, we intend the word "appearing" in its deepest sense, namely, we are making no distinction between "appearance" and "being": something that in *any* experiment appears to us as curved, i.e. can only be observed to be curved, *is* at all the effects curved. If for a moment we abstract our mind from the Universe as it really is, and try to think at a space-time without matter and light, we can think that this curvature manifests itself under the form of a cosmological constant. In any case, the observer feels

space has to be considered by looking at the string space in its whole (for more details and discussion, see for instance Ref. [11]).

a "ground" gravitational field fulfilling the entire space. Our question is then translated into understanding how light and matter come out from a space-time with gravitation (a cosmological constant).

Now we are in a position to understand this, namely how is it possible that a simple space-time "creates" light and matter: the physical system possesses a symmetry, according to which the three forms of energy appear as three aspects of the same phenomenon. The situation is somehow similar to that of electromagnetism: in that case, electric and magnetic fields are related by a Lorentz transformation, a transformation among the four extended space-time coordinates. Here the transformations are "S", "T", "U" dualities, symmetries acting of the full string space, not just on the ordinary space-time. There is therefore no "rest-frame" in the space-time where one of these dual aspects can be switched off. Under these transformations, what changes is the role of the degrees of freedom appearing in a certain construction: e.g. the gravity sector is mapped to the matter sector, or to the gauge sector, etc... Therefore, we can say that

a finite space-time possesses energy, a sort of "surface tension", for the simple fact of existing. This energy is equivalently due to "ground gravitation" (cosmological term), radiation and matter.

The equivalence of these three sectors is established at the level of the "orbifold" splitting of the string configuration into three sectors. In the next sections we will discuss how minimization of entropy requires to move a bit away from the orbifold point, in order to account for finer differences among the various degrees of freedom. These cannot be seen at the orbifold point, where too many moduli are frozen. This leads to different mass values for any kind of particles. Analogously, the non-exact symmetry between gauge and matter sectors tells us that the basic "$S - T - U$" symmetry is somehow softly broken. As a consequence, we can expect effects, driven by the moduli of the extended coordinates, and therefore of second order with respect to the energy densities, which distinguish the details of these sectors. Of this kind are the perturbations that lead to inhomogeneousness in space and time, such as galaxy clusters etc.

6 Masses and couplings

In section 3.2 we analyzed the structure of the minimal entropy string vacuum, $\psi^{\rm min}$, by considering a set of "overlapping" dual constructions. By inspecting the configuration, non-perturbative for any string construction, through these "slices", we could get a sufficiently complete insight into the structure of the "singularities" of this vacuum. This was enough in order to figure out what the structure of the minimal entropy configuration is. On the other hand, any dual realization is forcedly built as a perturbation around a zero value of a coupling. As we have inferred, in $\psi^{\rm min}$ only space time coordinates remain at the end extended: all the "internal" ones are of Planck size. The theory is therefore strongly coupled, and no one of the dual constructions could provide a *quantitative* estimate of whatever observable. Indeed, some of the properties could be investigated only by trading, in appropriate dual constructions, the space-time coordinates for the internal ones.

An exception was the energy density of the Universe, that could be non-perturbatively evaluated. In our scenario, its value is by definition just the mean value of the identity on the string vacuum, rescaled by an overall normalization. The latter is given by the Jacobian of the transformation of the space-time to the string/Planck scale. The result is therefore insensitive on the details of the vacuum itself. Any other quantity we may want to compute is on the other hand sensitive to the details of the solution $\psi^{\rm min}$, and requires to be approached in another way.

The way we will proceed can be called "semi-perturbative". Namely, we will consider the class of Z_2 orbifold constructions, and, in particular, the one of minimal entropy, as a set of "bare" configurations, something reminiscent of the asymptotic, free states at the ground of any field theory perturbative scattering computation. It is easy to realize that the configuration of minimal entropy cannot exactly lie at the orbifold point: this is in fact a point of partial symmetry restoration. In order to see this, simply consider that, in passing from the supersymmetric orbifold to the configuration of broken supersymmetry, the matter sectors get replicated into three copies that, at the orbifold point, appear as completely identical. This subtends a symmetry among these orbifold "planes", not unexpected because it is inherited from the symmetry of the orbifold projections. We already mentioned that nevertheless the Z_2 orbifold point could be considered a good choice for the "bare" configuration, because the phase space is extremely reduced (and therefore entropy too) being maximal the number of twisted moduli. We expect the real configuration to be somehow "slightly displaced" in the moduli space from the orbifold point. This "displacement" should be driven by the non-twisted moduli of the theory, namely those related to the space-time. In a vacuum, such as the one of minimal entropy, in which T-duality is broken, we expect that the corrections to the orbifold point, provided by switching on these moduli, are weak. Namely, they should be of the order of some (positive) power of the inverse of the age of the Universe, as compared to the size of the internal space (order one in reduced Planck units). For any observable A, we expect then:

$$< A > \; = \; < A >_{Z_2} + \Delta A \, , \tag{6.1}$$

where:

$$\Delta A / < A >_{Z_2} \; \sim \; \mathcal{O} \left(1 / \mathcal{T}^p \right) \, , \qquad p > 0 \, . \tag{6.2}$$

The requirement of entropy minimization, by imposing that the "fluctuations" around the point of maximal twisting be "minimal", allows to "keep under control" the corrections. We can then follow a kind of "perturbative approach", built around a non-perturbative string configuration: the minimal entropy orbifold point. In the following, we will make an extensive use of this idea, in order to investigate the mass of particles and fields constituting the "low-energy" spectrum of the theory.

6.1 Exact mass scales

We consider here the mass scales that, referring to stable states of the theory, can be non-perturbatively computed in an exact way. Exactness is here limited to the minimal entropy configuration: as we already pointed out, the functional 2.18 counts with a non-vanishing weight also non-minimal and non-critical configurations which, although suppressed, nevertheless contribute to the mean value of the observables of the Universe. In this section we will not be concerned with these effects.

6.1.1 The mean mass scale

We start by considering what is the "mean mass scale" of the Universe. This is defined as:

$$< m > = \sum_i < i \,|\, m \,|\, i > = \sum_i m_i P_i \,, \tag{6.3}$$

where $|\, i >$ are the mass eigenstates, with probability P_i, and the m_i the corresponding eigenvalues. It must therefore not be confused with the mass/energy density computed in the previous chapter; it is the mean eigenvalue of the mass operator (the Hamiltonian at rest, or, better, the ground step in the tower of energies of a field/particle). Unfortunately, for any finite value of the space-time volume, the theory is strongly coupled, and a perturbative approach to this computation, even approximated, is not possible. Any perturbative string construction is in fact based on a factorization of the string coordinates, that splits "additively" their contribution to the mass of the states. There are however good reasons to expect that, once non-perturbatively resummed, the coordinates (the compactification radii), should mix up "multiplicatively". In other words, while in any (forcedly perturbative) construction the coordinates contribute to the mass as:

$$M \sim \frac{m_1}{R_1} + \frac{m_2}{R_2} + \ldots \,, \tag{6.4}$$

non-perturbatively the radii should better enter as:

$$M \sim \left[\frac{m_1}{R_1} \times \frac{m_2}{R_2} \times \ldots \right]^{\frac{1}{\#\,\mathrm{of}\, R_i}} . \tag{6.5}$$

In order to understand this, we have to consider that all matter states, being intrinsically strongly coupled, are non-perturbative. We recall that the leptons arise as the singlets from the breaking of the $\mathbf{4} \to \mathbf{3} \oplus \mathbf{1}$, of a strongly coupled sector. Although they end up to

feel just weak interactions, they "live" on a sector of the vacuum which is non-perturbative with respect to the sector that gives origin to the weak interactions. There is no explicit construction in which both these sectors appear at the same time as weakly coupled. The mass eigenstates are therefore made of bi-charged states, that feel both the weakly and the strongly coupled sector of the fundamental theory. In order to "see" them as free, perturbative states, one should map to a dual picture in which a coupling of order 1 is mapped to a weak coupling. This is a logarithmic mapping $R \to X = \ln R$, under which the coupling $g \sim 1 \to \ln 1 = 0$. On the other hand, under this operation perturbative objects would become non-perturbative. For instance, weak interactions would not be anymore weak, unless we keep the "radius" of space-time sufficiently small. We can however think to ideally map to this picture at the beginning of the evolution, i.e. at the Planck scale. Once understood how things work, we can extrapolate the results also to the present time regime.

From this point of view, the orbifold constructions we described in section 3 correspond to a "linearized" representation of the vacuum; they are related to the actual, physical situation by some kind of exponential/logarithmic map. The matter sectors, as they appear in the three (four if we count both type IIB and IIA) dual constructions, correspond to the "logarithmic picture"; the real vacuum is obtained by exponentiation. In the language of Lie groups, this operation is what allows to pass from the algebra to the group: a direct sum is mapped to a direct product. In this light, we understand why in certain cases we obtain a fake duplication of bi-charged states, that appear as *different* states, charged with respect to perturbative groups arising from different sectors (e.g. the matter states in the bulk and D-branes sectors in the type I realization). Under exponentiation, instead of a sum we get a product, and the states duplicated in order to be charged under a <u>sum</u> of representations become states bi-charged under the <u>product</u> of groups. This transformation does not affect our previous computations of the energy densities: as we already said, the value of these densities simply accounts for the Jacobian of the transformation of the mean value of the identity from the string vacuum to the effective action in the Einstein's frame, and is therefore insensitive to the representation of the coordinates of the vacuum. Any such change is compensated by a change in the normalization. On the contrary, a mass expression consisting of the sum over the contributions of shifted coordinates turns out to correspond to a sum of logarithms of the coordinates of the "physical" picture:

$$m \sim \sum_i 1/X_i = \sum_i \log R_i \,. \tag{6.6}$$

Once pulled back, we obtain the multiplicative formula 6.5. Therefore, since mean values are computed as additive averages in perturbative (e.g. orbifold) string constructions, and these correspond to a "logarithmic" representation of the physical vacuum, we expect the mean value of the mass, as is computed in a perturbative string construction, to correspond to a logarithmic average over the coordinates of the "physical" space:

$$< m > \;\sim\; \sqrt[D]{\prod_i^D \frac{1}{R_i}} \,, \tag{6.7}$$

where D is the full dimension of space, internal and external. In fact, all the string excitations

are basically the (quantum) modes of expansion of the geometry of the target space. The mean mass, i.e. the mean ground energy level, is the value that, if attributed to all the states, sums up to give the total energy. It must therefore correspond to the inverse of the mean radius, intended as the value of radius such that, once attributed to all the coordinates, namely, when considering the string target space as a symmetric space, a "hypersphere", corresponds to the D-th root of the volume, as it is for a hypersphere. Therefore, in order to calculate it, we don't need to know much details about the minimal entropy string vacuum, except for:

1. the number of dimensions string theory lives in;

2. the volume of the target space.

We have seen that three space coordinates are extended up to $\mathcal{T}$, while the other ones are frozen at the Planck scale. Once this is taken into account, expression 6.7 reads:

$$< m > \;\sim\; \frac{1}{\mathcal{T}^{3/D}} \,. \tag{6.8}$$

In order to obtain the correct result, it is now just a matter of inserting the correct value of the full space dimensionality, D (here D is the dimension of space, so that the full space-time has dimension $D+1$. There is here a subtlety. Until now we have considered, as is usually done, *linearized* realizations of the string space. According to these, the string space appears as a 3+1 + $(D-3)$ dimensional space, with 4 extended and $(D-3)$ size-one, "internal" coordinates. We know however that from a non-perturbative point of view space-time is not so simply factorized. As we discussed, it seems that we are in the presence of 11 dimensional curved space-time, that, owing to the linearization introduced in the various dual "slices" of the theory, gives the impression to be twelve dimensional. Twelve dimensions are precisely what is required in order to embed in flat space an 11-dimensional space-time, of which 10 are coordinates of the curved space-like part. The "true", intrinsic space-time dimension is therefore 11, not 12, and the mean value of the mass scales as:

$$< m > \;\sim\; \left[\sqrt[10]{\left(\prod_i^{10} R_i = \mathcal{T}^3 \times 1^7 \right)} \right]^{-1} \;=\; \frac{1}{\mathcal{T}^{3/10}} \,. \tag{6.9}$$

If we include also the correct normalization of the mass, which, according to the Heisenberg's Uncertainty relation, $\Delta E \Delta t \geq \frac{1}{2}\hbar$, should be proportional to 1/2 the inverse of the space-time length, we conclude that the true value of the mean mass is:

$$< m > \;=\; \frac{1}{2}\, \mathcal{T}^{-\frac{3}{10}} \,. \tag{6.10}$$

In this expression, the time $\mathcal{T}$ is the age of the Universe as seen from the string frame. The age of the Universe, as derived with interpolations based on the usual Big Bang cosmology, is supposed to range from 11,5 and 14 billions years. Its central value is therefore $\sim 12,75 \times 10^9$ yrs, ($\sim\, 5 \,\times\, 10^{60} \mathrm{M}_{\mathrm{P}}^{-1}$, see Appendix A). Inserting it in 6.10, we obtain:

$$< m > \;\sim\; 7,49\, \mathrm{GeV} \,. \tag{6.11}$$

$\star\star\star$

In this section we have encountered for the first time a multiplicative behaviour of mean values, obtained as the non-perturbative resummation of an expression which, in a perturbative representation of the string vacuum, appears as a sum of terms. The full expression of mean values is obtained from the functional 2.31. As it is defined, it is indeed a sum over configurations ψ. Given an "operator" A, corresponding to a certain measurable quantity, its mean value is defined as:

$$< A > \; = \; \int \mathcal{D}\psi \, A \, \mathrm{e}^{-S} \, . \tag{6.12}$$

Considered as a value to be computed over all the configurations entering in the integral, the mean value is certainly a sum:

$$< A > \; = \; \int \mathcal{D}\psi \; < A_\psi > \; \approx \; \sum_\psi A_\psi \, . \tag{6.13}$$

However, any configuration ψ is actually a full, non-perturbative (string) vacuum. As we have seen, in general any such configuration ψ (for instance, on the minimal entropy configuration at a certain volume V, $\psi_V^{\rm min}$), is a product of systems, and its statistics follows the multiplicative laws of composite systems. For any fixed vacuum, the measure of the integral on the r.h.s. of 6.12 is a product, over all the states of $\psi_V^{\rm min}$, of their probability (raised to the probability itself). The multiplicative nature of mean values as a function of the coordinates of a single string configuration (such as for the minimal entropy, dominant configuration) is therefore related to the multiplicative nature of probabilities for a composite system. An additive correction shows out when considering the perturbations to the dominant behaviour, due to the non-vanishing contribution of non-minimal vacua.

Let's see how this works in the case of the mean mass. Similarly to what is usually done in the ordinary path integral, we can imagine to produce the insertion of A in expression 6.12 by switching on, in the exponential of entropy, currents J that couple to the operator. In the case of the mass, these are radii deformations. Since, according to 4.5, entropy scales as the square of the "radius" (i.e. the age of the Universe), we have:

$$\exp -S \; \approx \; \exp - \left(\prod^n R_i \right)^{2/n} \; \to \; \exp - \left(\prod^n (R_i + R_i J) \right)^{2/n} \, . \tag{6.14}$$

Notice that the integral deformation is $R_i \to R_i(1+J)$ and not $R_i \to R_i + J$: the latter would be an infinitesimal deformation on the tangent space. The mean mass is therefore given by:

$$< m > \; \approx \; \left[\frac{\delta}{\delta J} \ln \mathcal{Z} \right]_{J=0} \; \approx \; \left(\prod^n R_i \right)^{1/n} \, . \tag{6.15}$$

Had we instead used as deformation the one of the tangent space, $R_i + J$, we would have obtained for the mean mass the additive formula $m \sim \sum 1/R_i$, typical of the traditional perturbative string approach.

6.1.2 *The neutron mass*

We want now to discuss the physical meaning of the mean mass scale just considered. According to its definition, eq. 6.3, the contribution to the mean value should be provided by the asymptotic stable mass eigenstate(s) of the theory. These are not necessarily elementary mass/interaction eigenstates: in general they will be compounds. Usually, one thinks at the singlets of the strong interactions, because the theory is constructed as a perturbative vacuum around the zero value of the electromagnetic and weak couplings. Here however the situation is different: a finite, non-perturbative functional mass expression, valid at any value of the space-time volume, corresponds to a regime in which not only the strong interactions are non-perturbative, but also the electroweak interactions cannot be considered weak: the perturbative description of electro-weak interactions is an approximation, whose degree of accuracy increases with the age of the Universe [40]. The true free mass eigenstates are neutral to both strong and electroweak interactions. The mean value 6.3 corresponds therefore to the average value of the mass of stable matter in the Universe. We will see in the following sections how precisely the mass hierarchy allows to identify the sub-Planckian spectrum with the known elementary particles. The structure of particles and interactions arising in this scenario (section 3.3, see also Ref. [2]), is the same of the Standard Model, except for the absence of the Higgs sector.

As we will see, masses and gauge couplings scale as powers of the age of the Universe. However, the time dependence of gauge couplings is much milder than that of masses:

$$g \sim \frac{1}{\mathcal{T}^{1/p}}, \quad m \sim \frac{1}{\mathcal{T}^{1/q}}, \quad p \gg q. \tag{6.16}$$

Therefore, already outside of a close neighbourhood of the Planck scale, we rapidly fall into a regime in which the gravitational interaction is weak, while all other interactions are still strong. This is the regime of interest for our problem (at precisely the Planck scale the configuration becomes trivial). In this phase, the only state neutral under strong, electromagnetic and weak interactions, is a "composite" made out of a neutron-antineutron pair at rest, and their decay products, i.e. the proton-electron-neutrino/antiproton-positron-antineutrino system. At the "strong" limit of the weak coupling, family mixings can be neglected because one can assume that all heavier particles have rapidly decayed to the ground family. As it happens for stable matter, the decay probability of the neutron is compensated by an equal probability of the inverse process of neutrino capture, and the system is stable under weak interactions. It is invariant under charge reversal, and stable under electromagnetic interactions as well. This is the *only* singlet under all the above interactions, and therefore the only mass eigenstate at finite volume. At the present age of the Universe, the volume of space-time is anyway large enough to assure weakness of the electro-weak interactions. This composite is therefore not necessarily a "bound state", as it has presumably been at earlier times. We expect expression 6.10 to account for the mass of the "composite bound state", i.e. roughly twice as much as the mass of the neutron-

[40] The behaviour of these couplings will be discussed in sections 6.2.4 and 6.2.5.

antineutron pair. Therefore:

$$m_{\mathrm{n}} = \frac{1}{4} < m > = \frac{1}{8} \mathcal{T}^{-\frac{3}{10}} . \tag{6.17}$$

By inserting in 6.17 the current value for the age of the Universe, as obtained by extrapolating data of experimental observations within the theoretical framework of Big Bang cosmology, we obtain a value quite close to the neutron mass. Namely, from 6.11 and 6.17 we obtain:

$$m_{\mathrm{n}} \approx 937 \, \mathrm{MeV} , \tag{6.18}$$

quite in good agreement with the value experimentally measured $939, 56563 \pm 0, 00028$ MeV [58]. Unfortunately, the age of the Universe is not known with enough accuracy in order to test our formula: the only thing we can say is that our expression is *compatible* with the current experimental extrapolations. A more correct analysis would require a new derivation of the value of the age of the Universe completely *within our framework*. Anyway, owing to the degree of approximation applied to the usual computations, we expect the data about the age of the Universe, obtained by integrating the equations of motion for the expansion of the Universe, to catch at least the correct order of magnitude.

On the other hand, within our theoretical scheme we can reverse the argument, and take the mass of the neutron as the most precise measurement of the age of the Universe. In this case, we obtain as its actual value:

$$\mathcal{T}_0 = 12, 62028271 \times 10^9 \, \mathrm{yr} . \tag{6.19}$$

The fact that our mass formula gives as average mass the mass of the neutron is nicely in agreement with what we would expect from a Universe behaving as a black hole. According to the common astrophysical models, a black hole is in fact the step just following the "neutron star" phase of a star at the end of its life. Our considerations of above suggest that the Universe, as "seen from outside", can be roughly thought as a kind of neutron star at the point of transition to a black hole.

6.1.3 The apparent acceleration of the Universe

We are now in the position to come back to the issue of the apparent acceleration of the Universe. We have seen that the average mass of the stable matter scales with time as:

$$m \sim \mathcal{T}^{-3/10} . \tag{6.20}$$

If we take this mass as the reference for the atomic mass scale, we derive that the above behaviour induces an apparent shift in the frequencies of the light emitted at different distances from the observer, i.e. at different ages of the Universe, due to the different scale of the atomic energy levels:

$$\frac{\tilde{\nu}_1}{\tilde{\nu}_2} = \left(\frac{\mathcal{T}_2}{\mathcal{T}_1} \right)^{\frac{3}{10}} . \tag{6.21}$$

Once "subtracted" from the bare red-shift 5.20, this gives an apparent, effective red-shift $z_{\text{app.}}$:

$$1 + z_{\text{app.}} = \left(\frac{\nu_1}{\nu_2} \right)_{\text{observed}} = \left(\frac{\mathcal{T}_2}{\mathcal{T}_1} \right)^{\frac{7}{10}}, \tag{6.22}$$

as if the Universe were expanding with rate $\tilde{R} \sim \mathcal{T}^{7/10}$, normally expected for a matter dominated era.

At the base of what is considered an experimental evidence of the accelerated expansion of the Universe is the observed acceleration in the time variation of the red-shift effect. Here, this effect receives a different explanation, in terms of accelerated variation of ratios of mass scales, and therefore of observed emitted frequencies. On the other hand, the absence of a real acceleration allows to avoid some consequences that may look paradoxical. A real acceleration would in fact introduce an asymmetry in space, due to the choice of a preferred direction, implying also the possibility of finding out the "center" of the Universe, as an absolute, preferred point in the space. Experimental observations agree however on the fact that the Universe appears to be basically homogeneous in all directions [41]. This argument should not be confused with the "weak" breaking of Lorentz invariance we discussed in section 3.4, which has to be regarded as a "quantum fluctuation", and is responsible for some inhomogeneousness in the matter distribution, a second order effect. The choice of a preferred direction is a "first order effect".

An obvious remark is that, indeed, when talking of "accelerated expansion", one usually refers to the time variation of the overall scale factor of the space-like part of the Robertson-Walker metric of the Universe. The point is that, either this acceleration has a physical meaning in the ordinary, Newtonian sense, and therefore it can, at least in principle, be detected through a local experiment, such as for instance the measurement of some asymmetry in the gravitational force. Or, if instead it is a pure acceleration of the fundamental scale of the metric, something that doesn't affect local experiments, it is then equivalent to an accelerated variation of the global mass scale. In this second case, it is therefore precisely the phenomenon we are talking about in this section.

6.2 Non-exact mass scales

We consider now the evaluation of those quantities, such as the mass excitations corresponding to the elementary particles, that cannot be easily computed in an exact way. Solving the theory for the masses of the low energy degrees of freedom, i.e. the observed particles and fields, is not a task as "clean" as it was the computation of the cosmological constant: here we are looking for parameters that highly depend on the details of the configuration, not just on its global properties. The string vacuum corresponding to the dominant configuration is non-perturbative, strongly coupled. As we have seen, the spectrum of the matter states can anyway be investigated by mapping to a set of dual pictures, "logarithmic pictures", in which the internal coupling, of order 1, becomes of order 0, and it makes sense to use the

[41] This acceleration would be on the other hand so small as compared to the acceleration due to the gravitational field of the earth that a direct measurement is quite unlikely to be obtained.

tools of perturbative string theory. This mapping is subtended in any perturbative representation of the minimal vacuum. We would be tempted to use the same approach also for the computation of the masses of matter states. However, this procedure doesn't lead to non-trivial results. In the log-picture the couplings of electromagnetic and weak interactions are in fact, in general, not small at all. At present time, from the point of view of the log-picture these interactions are at the strong coupling: they are in fact given as logarithms of the moduli of the extended coordinates. The log-picture helps us therefore only at times close to the Planck scale, when all the coordinates are of order 1 in the ordinary picture. For the purpose of investigating the spectrum, i.e. basically counting the degrees of freedom, this was not a problem: the spectrum remains in fact the same at all the space-time scales, from a neighbourhood of the Planck scale up to a neighbourhood of infinity.

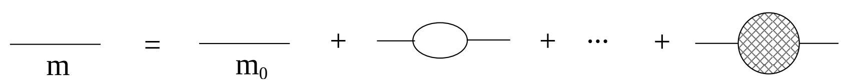
For what concerns their masses, things are rather different: the masses of light (sub-Planckian) excitations are themselves a "second order effect", they strongly depend on the moduli of the extended space-time. The idea of considering the masses computed at the Planck scale as the "bare" masses of our construction, to be corrected at later times of the evolution by inverse roots of the age of the Universe, doesn't work: we don't have a rationale to control these corrections. Simply imposing their agreement with the "initial conditions" at time 1 in Planck units is of no help: in the "exponential picture", the "true" vacuum, these corrections act multiplicatively (as does a resummation of an additive series in the logarithmic picture), and any multiplicative correction given by some power of the age of the Universe trivially reduces to 1 at the Planck scale. There is therefore no way to approach the problem of computing the masses of these states by using some kind of perturbation around a bare value, computed with traditional string methods, in some appropriate regime or dual picture. In order to evaluate the masses of the elementary particles we will therefore follow a completely different procedure, that we now illustrate.

6.2.1 Entropy and Mass

In section 3, when discussing the configuration of minimal entropy, we remarked that the orbifold description does not allow us to fully account for differentiations in the geometry of the internal string space that depend on the moduli related to the extended space-time. At the Z_2 orbifold point, all the internal coordinates look the same, and the shift along the space-time coordinates that gives origin to the mass of matter states acts in the same way on all the particles of the "low energy", i.e. sub-Planckian, spectrum. This is a consequence of the symmetry among the orbifold projections. As a result, leptons and quarks acquire all the same mass, and all the families have the same weight; this fact seems to imply the existence of an un-broken symmetry among all matter states. However, this is an artifact of the orbifold point, which freezes too many moduli, and oversimplifies the configuration. It is clear that minimization of entropy requires the breaking of this internal symmetry. The orbifold point corresponds to the extreme situation in which the breaking of this symmetry is so strong that we don't see the massive gauge bosons of the broken symmetry just because they have infinite mass. If we want to investigate the details of particle's masses, and distinguish particles and families, we have to go deeper through this point, and somehow

"switch on" the frozen moduli. The problem of computing masses of particles which are not singlets under all interactions is rather complicated, because it requires the evaluation of a mixture of perturbative and non-perturbative quantities: there is no point at which all interactions are weak. As we said, an evaluation in the logarithmic picture would require a correction at finite time, in order to obtain the present day value of masses. This correction would then be computed in a picture in which some interactions are strong.

In order to overcome these technical difficulties we will consider the problem from the point of view of the "thermodynamics" involved in the renormalization process. In order to understand the rationale underlying this procedure, let's consider the corrections to the mass of a particle, represented by an elementary state of the theory. As we said, the matter states appear as free, elementary states of the theory only in a logarithmic representation of the space, and only at a time close to the Planck scale. In the logarithmic picture, the corrections appear as the sum of a series of insertions in the free propagator:



(6.23)

The mass renormalization can be seen as a physical process in which the final state is the same particle as the initial state: no lines are propagating out of the bubble. It corresponds therefore to an adiabatic process. The first law of thermodynamics tells us that in this case,

$$dQ = 0 = dE - TdS \,. \tag{6.24}$$

If the temperature were constant, we would conclude that the variation of energy along this process corresponds to the variation of entropy. However, the temperature of the system is related to the energy, and varies with energy.

Consider the case of the Universe itself. We have seen that its mass, i.e. its rest energy, scales with time, $\mathrm{M}_{\mathrm{Univ.}} \sim \mathcal{T}$, while the temperature scales as the inverse of its age: $T \sim \mathcal{T}^{-1}$. Finally, entropy scales as $S \sim \mathcal{T}^2$. In this case, the relation 6.24 reads therefore:

$$dE \equiv d\mathrm{M}_{\mathrm{Univ.}} \sim \frac{1}{\mathcal{T}} \times dS(\sim \mathcal{T})\, d\mathcal{T} \approx d\mathcal{T} \,.$$

There is no surprise in the fact that here also a mass is associated to thermodynamics and therefore to entropy: we have to do with a quantum version of General Relativity, and the mass must enter in all the relations where energy is involved.

Let's now consider the case of a "universe" consisting of just one particle. In other words, let's restrict these considerations to the system illustrated in fig. 6.23. We have already discussed a similar situation when considering the Uncertainty Principle, in section 2.2. In this case, the "universe" behaves in itself like a quantum black hole. It possesses a

"temperature", which is related to the inverse of its "age", and is therefore proportional to its minimal energy gap, set by the (generalized) Schwarzschild relation 2.13. As discussed in section 2.2, the energy gap is the mass of the particle itself. Equation 6.24 reads therefore:

$$\frac{dm}{T} \sim \frac{dm}{m} = dS\,. \tag{6.25}$$

Once integrated, it gives:

$$\ln m = S + \text{const.}\,. \tag{6.26}$$

In order to evaluate entropy, we proceed as follows. We expect the probability of a process (such as a decay process) to be proportional to the volume of the phase space. For instance, the probability amplitude for a particle that can decay along, say, 10 channels, is twice as large as the amplitude for a particle that can decay only along 5 channels, provided all parameters are the same (strength of the couplings etc...). The key point is that, in our case, at its starting point the system does not possess a defined dynamics. It is not the probability that will be determined by the strength of the coupling, but the other way around: it is the size of the renormalization of the coupling that will be determined by the volume of the phase space of the process corresponding to that coupling. The effective coupling will in turn enter in the expression of the probability decay once the "effective theory" at a certain order will be determined. An effective theory, corresponding to a certain effective action, should be thought of as an "intermediate step": the ingredients of the effective theory are parameters computed up to a certain degree of approximation, that will be used in order to improve the accuracy of the approximation. If we think the string space as fibered over the "space-time" base, we can say that for any coordinate of the fiber there is a phase space of space-time size. Since the size of the internal coordinates, those of the fiber, is one in Planck units, the volume of the phase space will be:

$$\mathcal{V}_P = \mu^{\beta}\,, \tag{6.27}$$

where μ is the length (the volume) of space-time, and β a coefficient. For instance, if we start with a phase space of volume μ^{β_0} and act on the states with a Z_2 projection that halves the internal space, the un-projected states will have a phase space at disposal for their decays given by $\mu^{\beta_0/2}$. The probability density is proportional to the inverse of the volume 6.27:

$$P(x) = \frac{1}{\mathcal{V}_P}\,, \tag{6.28}$$

and entropy, defined as in 2.17, is then:

$$S = -\int_{\mathcal{V}_P} dx\, \frac{1}{\mathcal{V}_P} \ln \frac{1}{\mathcal{V}_P} = \beta \ln \mu\,, \tag{6.29}$$

where the results follows immediately from the fact that $\mathcal{V}_P$ is a constant in the domain of integration. After a projection like the one of above, entropy will be reduced by one half: $S \to (\beta/2) \ln \mu$. The mass renormalization reads:

$$\ln m = \beta_0 \ln \mu + \beta \ln \mu + \text{const.}\,, \tag{6.30}$$

that we can also write as:

$$\ln m \,=\, \ln m_0 \,+\, \beta \ln \mu \,+\, \mathrm{const.}. \tag{6.31}$$

This expression suggests that the term $\beta \ln \mu$ can be calculated in a logarithmic representation, and the coefficient β, corresponding to the "exponent" of the volume of the subspace of the phase-space of the particle, can be obtained by computing the amount of projections acting on the sector under consideration. We have separated a term β_0, because we must allow for a minimal value of entropy. At finite space-time volume, the minimal mass gap is not arbitrarily small. This sets also the minimal entropy allowed for the process, and results in a "minimal subtraction" in the phase-space volume. After exponentiation, we obtain:

$$m \,=\, (\mathrm{Const.}) \,\times\, m_0 \, \exp(\beta \ln \mu) \,=\, (\mathrm{Const.}) \,\times\, m_0 \, \mu^{\beta} \,. \tag{6.32}$$

The "bare" mass m_0 is given by the inverse square root of the age of the Universe: $m_0 = 1/\mathcal{T}^{1/2}$. This value, common to all particles, is produced by the shift acting along the space-time coordinates, that we described in section 3.2.4. In the logarithmic picture, where, owing to the linearization of space, shifts along the coordinates contribute additively, it appears as an additive term:

$$\ln m_0 \,=\, \frac{1}{2} \ln \frac{1}{\mathcal{T}} \,\equiv\, \beta_0 \ln \mu \,, \tag{6.33}$$

as in 6.31. In the two pictures the mass renormalization reads therefore respectively:

$$\begin{array}{lccl} (\log \text{picture}) & \tilde{m} \, [= \, \ln m] & = & \dfrac{1}{2} \ln \dfrac{1}{\mathcal{T}} \,+\, \beta \ln \mu \\[2ex] & & \equiv & \tilde{m}_0 \,+\, \beta \ln \mu \\[2ex] & \xrightarrow{\exp} & & \\[2ex] (\text{real picture}) & m & = & \mathcal{T}^{-\frac{1}{2}} \,\times\, \mu^{\beta} \,. \end{array} \tag{6.34}$$

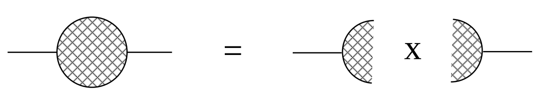
In these expressions, μ is the age of the Universe raised to some power: $\mu = \mathcal{T}^p$. It is clear that any change in p can be reabsorbed by a change in β and β_0. We set by convention $\mu \equiv \mathcal{T}$, so that $\beta_0 = -1/2$. By inserting these values in 6.32, we obtain that masses scale as:

$$m \,=\, (\mathrm{Const}) \times \mathcal{T}^{\beta - 1/2} \,, \tag{6.35}$$

The relation between mass and volume of the phase space tells us that, the more are the decay channels of a particle, the larger is its entropy and the correction to the mass (all this is illustrated in figure 8 of page 67). Heavier particles possess a huger decay phase space: quarks are heavier than leptons, and among leptons neutrinos are the lightest particles. Inside each family of particles, the heavier (for instance the top as compared to the bottom of an $SU(2)$ doublet) has the larger absolute value of the electroweak charge. In each family,

the lightest particle is the one which has less interactions, or less charge (and therefore a lower interaction probability). For instance, $|Q_\nu| < |Q_e|$, $|Q_b = -1/3| < |Q_t = +2/3|$, and quarks, that feel also the $SU(3)$ interactions, are heavier than leptons. The reader may point out that, as a matter of fact, there are evident exceptions to this rule. For instance, the case of the first quark family: the down quark is heavier, although less charged, than the up quark. Once again, we are faced to the fact that the separation of space-time and phase-space and the classification of elementary excitations as it appears in the orbifold approach is too schematic. We will come back to the discussion of these details in section 7.1.2; here we limit the discussion to the general behaviour. Along this line, we can view the lightest particle as the end-point of a chain of projections that reduce the symmetries of the internal space. It corresponds therefore to the less entropic step, and, as expected, is the less interacting one.

In the traditional thermodynamics and statistical physics, entropy is related to either the macroscopic, or the microscopical, description of the phase space, where a key role is played by energy. Here we are extending the relation to include also the mass. This doesn't come out as a surprise for a general relativistic extension of these quantities, in a theory which includes in its domain also black holes (among which, in particular, the Universe itself!). The mass is the ground energy of a system, and corresponds to a "ground entropy". An object with mass possesses a "ground entropy". In the case of a black hole, entropy can be seen as somehow related to a counting of the elementary-state paths going from outside into the horizon of the black hole [59, 60, 61, 62]. In a similar way, here the relation between mass corrections and entropy, established in 6.23 and 6.24, suggests that the "ground entropy" of a particle can be seen as a way of counting the paths leading to the particle. The "bubbles" of the terms in 6.23 represent in fact paths that come out of the particle and end up back into the particle itself. The equivalence with the usual interpretation of the entropy of a black hole is more evident if we ideally view these terms as made up of two mirror cuts:



(6.36)

Each half looks like a decay process, or the other way around as the process of formation of a black hole. Notice that, although the mass of a particle decreases with time, the "ground entropy" increases. This is due to the fact that entropy is related to the statistics of the physical system, i.e. to the structure of the phase space. As it can be seen from the Uncertainty Relations ($\Delta E \Delta t \geq 1/2$), in this space the volume of the minimal cell goes like $\sim 1/\mathcal{T}$. It decreases therefore much faster than any mass scale, which is bounded by the square root scale: $m \geq 1/\mathcal{T}^{1/2}$ for any mass m. The "ground" phase space volume of a particle is:

$$\mathcal{V}_0 = \frac{m}{\Delta E} \rightsquigarrow \frac{\mathcal{T}}{\mathcal{T}^{\frac{1}{2}}} \sim \mathcal{T}^{1/2} \,. \tag{6.37}$$

As time goes by, the phase space of a particle increases, and therefore also its entropy. In

section 6.2.11 we will comment on the relation of this approach to the one based on the "geometric probability" techniques.

6.2.2 *Running of couplings and mass ratios*

In the previous section we encountered, as a basic ingredient of mass formulae, the expression μ^β. This corresponds, in the logarithmic picture, to $\beta \ln \mu$. We are used to see this as the typical one-loop renormalization of the inverse of a gauge coupling: $\frac{4\pi}{g^2} \sim \frac{4\pi}{g_0^2} + \beta \ln \mu$. Indeed, the idea of renormalization group is based on the possibility of running the parameters as functions of a scale, in such a way that the sum of higher order terms can be re-written so as to reproduce the functional expression of the first non-trivial correction. This means that there is always an appropriate scale $\tilde{\mu}$ and an effective "beta-function" $\tilde{\beta}$ for which $\tilde{\beta} \ln \tilde{\mu}$ constitutes a good approximation to the expression of the coupling, at least for its perturbative part. Since the logarithmic picture is for us precisely the representation in which the coupling of the theory vanishes, this could be taken somehow as the definition of the running of a coupling in this picture. And indeed, a perturbative series such as:

$$\partial \ln \alpha / \partial \ln \mu \approx \sum_n \beta_n \alpha^n \,, \tag{6.38}$$

somehow looks like the expansion around a small value of the coupling of an exponential expression, of the type:

$$\alpha(\mu) \sim \exp\left[\beta \ln \mu\right] \,, \tag{6.39}$$

over $\ln \alpha$:

$$\ln \alpha \sim \beta \ln \mu \,. \tag{6.40}$$

This suggests that expressions like 6.32 and 6.34 involve the non-perturbative, resummed values of couplings of the theory. In particular, the ratios of masses of different particles would be:

$$\frac{m_i}{m_j} = \frac{\mu^{\beta_i}}{\mu^{\beta_j}} = \frac{\alpha_j}{\alpha_i} \,, \tag{6.41}$$

i.e. they would be given by ratios of couplings, in turn proportional to the inverse of the volumes occupied by the particles in the phase space. How can we understand this? First of all, we must consider that, whenever two particles have the same mass (in particular, a vanishing mass) and the same transformation properties, there is in the theory a degeneracy corresponding to a symmetry. In this work we have analysed the vacuum with the language of orbifolds, and arrived to the replication of matter into three families, as the natural consequence of the separation of the string space, in the linearized picture suitable for the orbifold representation, into three planes with the same properties. The equivalence of these planes reflects the interchangeability of the corresponding orbifold projections. As we remarked, this linearization of the space, or better this simplification, in which a smooth curvature appears "summarized" at the orbifold fixed points, constitutes a good approximation, because it corresponds, up to corrections of the order of some roots of the inverse space-time length, to the configuration of minimal entropy. However, it remains an approximation. In particular, it doesn't allow us to follow the fate of the moduli of the symmetry

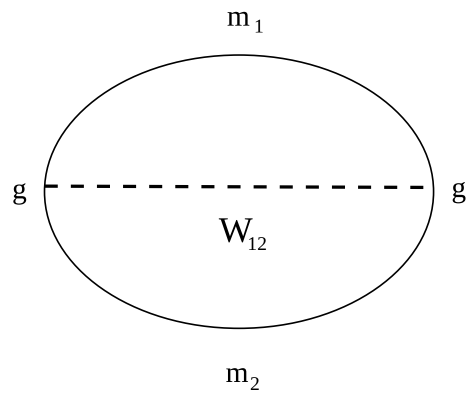
among the orbifold planes. The vacuum appears to us as a "frozen", rigid configuration with an apparent discrete symmetry. We don't see the corresponding gauge bosons: they have infinite mass, while the mass of the matter states vanishes. On the other hand, we know that in any perturbative realization of a string vacuum with gauge bosons and matter states charged under a symmetry group, gauge and matter originate from T-dual sectors. For instance, in heterotic realizations the gauge bosons transforming in the adjoint of the group originate from the currents, while the fermions transforming in the fundamental representation originate from a twisted sector. We can figure out what happens slightly away from the orbifold point, if we consider that a non-freely acting orbifold, consisting of only twists, can be seen as a singular case, obtained at a corner of the moduli space, of a more general freely acting projection. In this case the effect would be to produce a non-vanishing mass for the matter states as the effect of a shift on the windings, and for the projected gauge bosons as a consequence of a shift on the momenta. At the twist orbifold point, the two kinds of masses are sent the one to zero and the other to infinity. Outside of this limit, the mass of the projected matter states scales in a T-dual way to that of the gauge bosons [42].

If we go to a "second order" in our approximation, and think of blowing up the neglected moduli, we must figure out a situation in which, starting from an initial maximal symmetry of the string space, we progressively reduce this degeneracy by introducing projections that generate a non-vanishing mass both for matter and gauge states. The mass shift for the gauge bosons is then approximately T-dual to the one of the matter states. In the breaking of the internal symmetry into a set of separate families of particles, matter acquires a light mass, below the Planck scale, while the gauge bosons of the broken symmetry acquire a mass above the Planck scale.

Let's consider the following diagram for the vacuum renormalization in the case of a

[42] In Ref. [16] (and more in general Ref. [11]) we discussed string constructions with $\mathcal{N}_4 = 2$ supersymmetry and $N_V = N_H$, in which the bosons of the gauge group (more precisely the vector multiplets) and the matter states (the hypermultiplets) were realized both on the currents, and transformed in the same representation of the gauge group. This structure was preserved by the action of rank-reducing orbifold projections. In these cases, the matter states were not T-dual to the gauge bosons, and the shift giving mass to the bosons gave the same mass also to the matter states. However, the realization of this situation was very peculiar and, as discussed in Ref [16], was based on freely-acting orbifold operations. The minimal entropy configuration doesn't belong to the class of these constructions.

two-particle theory with a broken symmetry relating particle 1 and particle 2:



(6.42)

This has to be thought not as a field theory diagram, but as a diagram for our effective non-perturbative theory. There is no question about non-renormalizability of a theory without a Higgs particle: here masses have to be treated as effective parameters, more or less like what one does with the electric field in a semiclassical approximation of a quantum relativistic scattering by an external field. The theory is finite, and we know from above that this diagram, intended as the non-perturbative resummation of the terms corresponding to this transition process, should correct the vacuum amplitude *multiplicatively*, by a factor 1 [43]: this means in fact that there is no renormalization of the vacuum, when for g, m_1, m_2, $m_{W_{12}}$ we take the "on-shell", non-perturbatively resummed physical parameters. By computing the contribution of this diagram, we find:

$$1 \approx g^2 \times \int \frac{d^4 p_1 d^4 p_2}{(\not{p}_1 + m_1)(\not{p}_2 + m_2)\left((p + p_1 - p_2)^2 - M_W^2\right)} \sim g^2 \int^{<p>} \frac{d^8 p}{p^4} , \tag{6.43}$$

where the integration has to be performed up to a cut-off $< p >$. This corresponds to the typical mass scale of the (sub)space in which the process takes place. In practice, we can consider that the subsystem consisting of the particles 1, 2 and the boson W_{12} "lives" in a space with "temperature" given by the average of the temperatures of the three states. Equivalently, we can think that the typical length of the process is the (multiplicative) average of the typical lengths associated to the three states. In any case,

$$< p > \sim \left(m_1 m_2 M_W^2 \right)^{1/4} . \tag{6.44}$$

Therefore, the integral pops out a factor $[< p >]^4 = m_1 m_2 M_W^2$, where the W_{12} mass is T-dual to the lightest particle mass:

$$M_W^2 = \frac{1}{m_2^2} , \tag{6.45}$$

where by convention we have chosen m_2 to be the lower mass, and T-duality is performed with respect to the Planck mass. From this we derive that:

$$g^2 m_1 m_2 M_W^2 = g^2 \frac{m_1}{m_2} \approx 1 \tag{6.46}$$

[43]We recall that for us the additive representation corresponds to a logarithmic representation of the physical string vacuum.

The ratio of the two masses is therefore given by the inverse square coupling of the broken symmetry. We stress that these arguments make only sense in our scenario, not in a generic field theory.

Let's now consider what happens when we act with more projections, and break further the symmetry of the string space. We start with an initial symmetry group $\mathbf{G}$ and break it into a product of subgroups: $\mathbf{G} = \mathbf{G_1} \otimes \mathbf{G_2} \otimes \ldots \otimes \mathbf{G_n}$. In the usual renormalization group approach we work in the algebra $\mathcal{G}$ of the group, where $\mathcal{G} = \mathcal{G}_\mathbf{1} \oplus \mathcal{G}_\mathbf{2} \oplus \ldots \oplus \mathcal{G}_\mathbf{n}$. The (inverse) effective coupling seems to renormalize additively: for instance, the one-loop beta-function of $SU(N)$ with gauge bosons in the adjoint and (massless) matter states in the fundamental representation is one half of the beta-function of $SU(2N)$ with an analogous content of matter and gauge states. Indeed, from the point of view of our approach, what seems to behave additively is just the logarithmic derivative of the coupling. On the other hand, the effective coupling determines the probability amplitudes:

$$P(A \to B) \sim \alpha_{(AB)} \, , \tag{6.47}$$

where, as usual, $\alpha_{(AB)} \equiv g^2_{(AB)}/4\pi$. A composite transition/decay process: $A \to B \to C$, corresponds to a rotation with an element of the group $\mathbf{G}_{(AC)}$, given by a product $\mathbf{G}_{(AC)} = \mathbf{G}_{(AB)} \times \mathbf{G}_{(BC)}$. This transition corresponds therefore to: 1) first a rotation with an element of the group $\mathbf{G}_{(AB)}$ and then: 2) a rotation with an element of $\mathbf{G}_{(BC)}$. Therefore, we expect the probability of the decay $A \to C$ to be the product of the decay probability of $A \to B$ and of $B \to C$:

$$P(A \to C) = P(A \to B) \times P(B \to C) \, . \tag{6.48}$$

In usual field theory computations, the effective coupling of a composite process is not simply given by the product of the effective couplings of the single parts: at any step of the process, we have other "decay products", such as boson propagator lines that either circulate in loops or come out of the diagram. Here however we are not interested in transitions taking place at the "field theory scale". The bosons corresponding to these broken symmetries have a mass above the Planck scale, and this process doesn't appear as a usual decay but rather as a "transition". The symmetry appears to be always broken at any sub-Planckian scale. Indeed, only when the boson is below the Planck scale we have to deal with a true decay, otherwise, the boson has to be considered as an "external field". When the boson mass is above the Planck scale, the transition looks more like a rigid symmetry. In this case, we don't observe any boson propagation; when the transition is "off-diagonal" with respect to the traditional particle classification into families, we interpret this phenomenon as a "family mixing", whose effect is collected in the off-diagonal entries of the CKM mass matrix. We will come back to these issues: in section 7.6 we will discuss the case of "field theory boson masses", i.e. those of the ordinary Z and W bosons of the weak interactions, and in section 8 (8.1) the non-field theory transitions, in particular their appearance as "rigid" transitions among particles. In the case of interest for us now, namely when the boson mass, T-dual to the mass of the particle, is higher than the Planck scale, there is no boson line really coming out from the line of the particle transition, and the meaning of internal boson propagators is the one discussed above, diagram 6.42 and expression 6.46. This process strictly follows the rule of composite probabilities as in 6.48, and the effective coupling for the transition

from A to C is given by the product of the effective couplings of the single steps:

$$\alpha_{(AC)} \propto \alpha_{(AB)} \times \alpha_{(BC)} . \tag{6.49}$$

The square coupling of the group $\mathbf{G} = \mathbf{G_1} \otimes \mathbf{G_2} \otimes \ldots \otimes \mathbf{G_n}$ is therefore:

$$\alpha_{\mathbf{G}} = \alpha_{\mathbf{G_1}} \times \alpha_{\mathbf{G_2}} \times \cdots \times \alpha_{\mathbf{G_n}} . \tag{6.50}$$

Putting together 6.46 and 6.50 we can now understand why the series of the masses of the elementary particles found in nature appears to roughly arrange in a logarithmic sequence: this ordering reflects the sequence of the couplings of the symmetries broken when distinguishing the various matter states. The couplings of the subgroups are proportional to an inverse power of the age of the Universe [44]:

$$\alpha_i \propto \mathcal{T}^{-\beta_i} , \tag{6.51}$$

and the ratios of masses are therefore given, as anticipated in Ref. [2], in terms of powers of the age of the Universe:

$$\frac{m_i}{m_j} = \frac{\mathcal{T}^{\beta_j}}{\mathcal{T}^{\beta_i}} = \mathcal{T}^{\beta_j - \beta_i} . \tag{6.52}$$

Notice that couplings are here proportional to a volume to a certain power, i.e. to a certain typical length, as we should expect from a compactified string theory. They, as well as masses, naturally "unify" at the Planck scale, $\mathcal{T} \to 1$. To be more precise, what are expected to unify at the Planck scale are the masses of $SU(3)$, $SU(2)_{\text{w.i.}}$ and $U(1)_{\text{e.m.}}$ singlets: in this limit, all these interactions are in fact strong. Indeed, at this scale, such elementary excitations may not even exist, in that the only possible state is the "totally strongly coupled" singlet described in section 6.1.2. The condition of unification has therefore to be taken here as only indicative of an asymptotic behaviour, rather than a statement of a real physical condition at the Planck scale. As we will discuss more in detail in the next sections, this condition implies that, in the case of quarks, what enters in 6.52 is the mass of a compound of three quarks, therefore three times the mass of the single free quark. More in general, we will have to distinguish the cases of neutral and electrically charged particles, and how the masses of single states are disentangled from those of $SU(2)$ singlets. Being only "virtual", the asymptotic behaviour, used in order to fix the relative normalization of the masses, is however of no help in order to fix the overall normalization.

In order to obtain the masses, we must first obtain the "beta-functions" β_i, β_j. According to our discussion, we cannot compute them using the rules of ordinary field theory: here we are interested in the full, non-perturbative beta functions. We can proceed as follows: we can determine the ratios of these beta-functions if we know the ratios of the phase-space volumes. For instance, if a projection reduces by one half the phase space, the beta-function

[44] We recall that, up to a redefinition of the exponent, the scale μ in eq. 6.39 and following, can be identified with the age of the Universe $\mathcal{T}$, as discussed after equation 6.34.

will be one half of the initial one. On the other hand, the phase space volumes can be determined if we know the spectrum of interactions of the various particles (i.e., the pattern of figure 8), but the important point is that, in this framework, this in principle is equivalent to knowing the chain of projections, $\equiv$ symmetry reductions, giving origin to the sector a certain massive state belongs to. In previous sections we have investigated the pattern of the projections by collecting information provided by patching dual perturbative constructions of the vacuum. This seems to suggest that even in this case the right picture in which to approach the problem is a logarithmic representation of the space. However, here things are more complicated. First of all, we are interested in a "second order" effect, driven by moduli frozen at orbifold point, and we already pointed out that, at present time, weak couplings are not small in this picture. In principle, since we are interested in the evaluation of constant coefficients, this doesn't constitute a problem: we can think to determine the value of these parameters at an early time, and then let masses run up to the age of interest.

Once the functional dependence of masses on the time is known, this goes straightforwardly. However, the problem is precisely in the kind of approximations we introduce in the explicit realization of the vacuum, i.e. in the linearization we apply to the string space, in order to represent it perturbatively (through orbifolds). In order to recover the dependence on the frozen degrees of freedom, we will have to combine several considerations in subsequent steps, that somehow constitute a kind of "perturbative approach", in which the fine details of the structure of the physical minimal entropy configuration are the better and better investigated. Anyway, this way of proceeding makes sense, because these corrections, depending on inverse powers (roots) of the age of the Universe, are at present time "small". Going through these steps will be the matter of the next sections.

Once obtained the ratios of beta-functions, in order to get all their values we must fix one of them. To this purpose, we must consider that the mass of the state with maximal, unbroken symmetry, does not change with time, it is a constant. Maximal symmetry, and therefore also supersymmetry, implies in fact that masses either vanish (as also the cosmological constant does), or they do not renormalize out of the initial value at the Planck scale: among the preserved symmetries, there is in fact also time reversal, so that masses do not run. In this case, we have:

$$m_{\mathrm{max.\,symm.}} \;=\; \frac{1}{2}\,\mathcal{T}^{\beta_{\max}-\frac{1}{2}} \;=\; \mathrm{Constant}\,, \tag{6.53}$$

where we have set the normalization of the mass as a function of the inverse of a proper time to be 1/2, as according to the Heisenberg's Uncertainty Principle. The condition 6.53 implies $\beta_{\max} = \frac{1}{2}$. With this normalization, at the Planck time the maximal mass is the one of a black hole, in agreement with our discussion of section 5, namely it is given by the relation: $2M = R$, with the identification $R = \mathcal{T}$. Therefore, at the Planck time the minimal mass excitation is forced by the Uncertainty Principle to be 1/2. On the other hand, being the Universe a Planck size black hole, this is also the maximal allowed mass excitation. If we allow one coefficient to be greater than 1, there exists a certain size of the Universe, *larger* than the Planck length, at which the mass is larger than the black hole's Schwarzschild mass. If we run back in time, starting from the present age, where this state appears among those of the sub-Planckian spectrum, this excitation drops out

from the spectrum, by definition formed by the degrees of freedom which are below the black hole threshold, *before* reaching the Planck scale. Therefore, it cannot correspond to a perturbation over the massless spectrum driven by the moduli of the extended space-time (those giving rise to the effective, "low energy" theory). From 6.52 we see that the overall normalization is the same for all the states which are $SU(3)$, $SU(2)_{\rm w.i.}$ and $U(1)_{\rm e.m.}$ singlets:

$$m_i \; = \; \kappa_i \mathcal{T}^{\beta_i} \, , \qquad \kappa_i = \kappa \;\; \forall \, i \, . \tag{6.54}$$

Therefore all these masses are expressed as:

$$m \; = \; \frac{1}{2} \, \mathcal{T}^{\beta - \frac{1}{2}} \; , \qquad\quad \beta \in \{ 0, \frac{1}{2} \} \, . \tag{6.55}$$

6.2.3 Elementary masses

A look at the masses of charged leptons reveals that, in first approximation, their ratios satisfy the relation: $(m_\mu / m_e) \sim (m_\tau / m_\mu)^2$. Similar, although not completely regular, ratios appear to relate also the masses of the other particles: up-to-up or down-to-down quark relations. This suggests the existence of a logarithmic ordering in the sequence of masses, that should be at the base of mass relations, in passing from one family to the other one. This structure would then be partially spoiled by the details, i.e. the peculiarities, of the "history" of each particle. We will now see how the entropy approach introduced in this work proves to be appropriate in explaining the various mass relations among elementary particles. Indeed, a logarithmic sequence comes out quite naturally if we consider the heavier particles as occupying a larger volume in the phase space: the structure relating the mass ratios is then directly related to the various volumes occupied by the particles. The logarithmic ordering reflects the sequence of symmetry reductions.

Let's indicate with p the "beta" function coefficient, i.e. the exponent of the phase-space volume, $\mathcal{V} = \mu^\beta$, of the maximally symmetric configuration, and with q the one of the end-point, minimal symmetry (minimal entropy) configuration. The couplings of the various subgroups in which the initial symmetry group gets divided are then given by:

$$\alpha_i^{-1} \; = \; \mu^{\frac{n_i - n_{i+1}}{p}} \, , \tag{6.56}$$

where the ratios $\frac{n_i}{p}$, with $n_i : p = n_0 \geq n_{\rm i} \geq q$, are the volume exponents of the various steps. n_0 corresponds to the last step (the one which selects the lightest neutrino). An accurate evaluation of masses is then equivalent to an exact determination of these coefficients. As we discussed, the dynamics of the Universe is "determined" by an entropy principle, entirely contained in 2.18. In practice, from a concrete point of view it is convenient to "approximate" the solution through subsequent steps [45]. Our "perturbative" approach starts with a first degree of approximation, consisting of a "rough" determination of the volume of the phase space of each elementary particle, as seen "at the Planck scale". This allows us to

[45] Precisely because of the fact that entropy determines the entire dynamics, a detailed evaluation of the coefficients is equivalent to a detailed evaluation of all the decay processes and their probabilities, for each particle.

map to the "logarithmic picture", where all couplings are perturbative, because they are of order one in the original picture. Further corrections of the weak coupling scale are to be expected at the present age of the Universe. This leads us out of the domain of a logarithmic picture; the problem can in principle be treated as an ordinary perturbative correction, whose complete evaluation must take into account the details of every decay channel. In any case, these corrections should be of second order, with a relative magnitude proportional to the inverse ratio of the phase volume of the particle under consideration and the one of its decay products. Namely, we expect:

$$m \,=\, m_0 \,+\, \delta m \,, \tag{6.57}$$

where

$$\frac{\delta m}{m} \,\sim\, \mathcal{O}\left[m_{\rm final} / m_{\rm initial} (\equiv m) \right] \,. \tag{6.58}$$

We will consider these corrections in sections 7.2–7.5.2.

Let's consider the first step of the analysis, the one in which we can map to the logarithmic picture. If the hierarchy of the three families were ordered in such a way that all the particles of the higher family were heavier than those of the neighbouring lighter family, we could immediately conclude that the phase space is divided by the particle's families into three parts, with volumes staying in ratios given by 3:2:1. Namely, in passing from one family to the other one, the phase space volume $\mathcal{V}$ should undergo a contraction $\mathcal{V} \to \mathcal{V}^{\frac{2}{3}} \to \mathcal{V}^{\frac{1}{3}}$. However, this sequence is valid only as long as *all* the particles of the "heavier" family are indeed all heavier than the particles of the following family. Otherwise, the pattern changes. In order to understand the implications of this condition, we must consider that relating the mass to the corresponding volume in the phase space roughly means that heavier particles are also the more interacting ones. On the other hand, saying that all the particles of the heaviest family are heavier than those of the lighter families means in particular that even the tau-neutrino, ν_τ, is heavier than the charm and strange quarks, something not expected for a particle definitely less interacting. Indeed, we expect the three neutrinos to be the lightest among all particles. This means that the phase space is *first* separated into the block of neutral and charged particles, and *then* each block is contracted according to the 3:2:1 rule. Apart from the separation of the $SU(4)$ symmetry into $\mathbf{1} \oplus \mathbf{3}$ of leptons and quarks, all the other separations are built on $SU(2)$ steps, as a consequence of the Z_2 structure of the projections. We expect therefore the $SU(2)$ coupling to play a key role in the mass ratios. In principle, an $SU(2)$ coupling factor separates also the up and down, both lepton and quark, inside each family. However, the coupling, and the group, of interest for us is not the left-moving $SU(2)$ of the weak interactions, which at the string scale remains unbroken : mass separations are in relation with the breaking of the $SU(2)_{(L)} \leftrightarrow SU(2)_{(R)}$ symmetry. Requiring that all neutrinos are lighter than any charged particle implies a "flip" in the action of the up/down separation of the $SU(2)$ doublets, such that the minimal block, the $SU(2)$-coupling, instead of acting "diagonally", by separating the tau lepton from its neutrino, acts off-diagonally, being interposed from the ν_τ and the lightest charged particle, the electron.

Let's start by considering the lightest steps: the mass ratios of neutrinos. They should be separated by "minimal blocks" consisting of $SU(2)$ coupling factors:

$$\frac{m_{\nu_\tau}}{m_{\nu_\mu}} \sim \frac{m_{\nu_\mu}}{m_{\nu_e}} \sim x\,, \qquad \frac{m_{\nu_\tau}}{m_{\nu_e}} \sim x^2\,, \qquad x = \alpha^{-1}_{SU(2)}\,. \tag{6.59}$$

According to the hypothesis of the 3:2:1 separation of the neutrino subspace of the phase space, a $x = \alpha^{-1}_{SU(2)}$ factor should also separate the mass of the lightest neutrino, $\nu_{\rm e}$, from the pure "vacuum" of the mass sector, namely the square-root energy scale corresponding to the basic shift: $\mathcal{T}^{-1/2}/2$, the scale which, in the language of 6.53, corresponds to $\beta = 0$. A further $\alpha^{-1}_{SU(2)}$ should then separate the heaviest neutrino from the lightest charged particle, the electron:

$$\frac{m_e}{m_{\nu_\tau}} \sim \alpha^{-1}_{SU(2)}\,. \tag{6.60}$$

As we discussed, from now on we should expect a 3:2:1 relation between the phase space sub-volumes of the three families in the logarithmic picture:

$$\ln \mathcal{V}(t,b,\tau)\,:\,\ln \mathcal{V}(c,s,\mu)\,:\,\ln \mathcal{V}(u,d,e) \quad\approx\quad 3:2:1\,. \tag{6.61}$$

In order to determine one of these volumes, what we need now is to know the down-quark-to-lepton mass ratio. From the down to the up quark the separation should be again an $SU(2)$ coupling factor. There are however subtleties. First of all, we must notice that it is $\mathcal{V}(t,b,\tau)$ that contains these factors at their "ground" level. The reason is that, as we have seen in section 3.3, the shift which, acting along the space-time coordinates, realizes a further "$SU(2)$ step" in the reduction of entropy, breaking the group of the weak interactions, also introduces a mass differentiation in the matter states of the same order of the scale of the breaking of the gauge group. The full width of the phase space separation corresponds therefore to the scale of the breaking, i.e. the heaviest scale at which a mass separation between matter states related to this broken symmetry appears. We will derive $\mathcal{V}(c,s,\mu)$ and $\mathcal{V}(u,d,e)$ as fractional powers:

$$\mathcal{V}(c,s,\mu) \sim [\mathcal{V}(t,b,\tau)]^{\frac{2}{3}}\;; \qquad \mathcal{V}(u,d,e) \sim [\mathcal{V}(t,b,\tau)]^{\frac{1}{3}}\;, \tag{6.62}$$

A second subtlety is that, in the case of quarks, we expect the $\alpha_{SU(2)}$ factor between the top and bottom quark to separate not the masses of the single quarks, but $SU(3)$ triplets, i.e. singlets of the confining symmetry. In other words, the equivalence of leptons and quarks is established at the level of asymptotic states, which are singlets for the confining theory. We expect therefore a factor 1/3 correcting the mass ratio between the top and bottom quark. This normalization is due to the requirement that, once run back to the Planck scale, only $SU(3)$ singlets (in this case quark triplets) go to the same limit mass value as the leptons [46]. The top-to-bottom mass ratio should therefore be:

$$\frac{m_t}{m_b} \sim \frac{1}{3}\,\alpha^{-1}_{SU(2)}\,. \tag{6.63}$$

[46]We will discuss below how further normalization factors are needed in order to take into account the fact that at the Planck scale also the electromagnetic and weak interactions are "strong", so that masses must be normalized with respect to singlets of these interactions too.

The mass separation between quarks and leptons is provided by a "Wilson line" which, as we discussed, divides the **4** of each family into $\mathbf{3} \oplus \mathbf{1}$. This separates the phase space in two parts of unequal volumes. In first approximation, this separation corresponds to disentangling "one quarter" of $SU(4)$, and therefore we expect the "up" of the **1** part to lie a $\sqrt{\alpha_{SU(2)}}$ factor below the "down" of the **3** part. This is the separation factor between bottom quark and τ lepton:

$$\frac{m_b}{m_\tau} \approx \frac{1}{3} \frac{1}{\sqrt{\alpha_{SU(2)}}} \,. \tag{6.64}$$

Again a 1/3 factor is needed in order to account for the passage from $SU(3)$ singlets to free quarks. Altogether, the top-tau mass ratio is:

$$\frac{m_t}{m_\tau} = \frac{m_t}{m_b} \times \frac{m_b}{m_\tau} \sim \frac{1}{3} \, \alpha^{-1}_{SU(2)} \times \frac{1}{3} \frac{1}{\sqrt{\alpha_{SU(2)}}} \,. \tag{6.65}$$

According to 6.62, the analogous separation for the first family should read:

$$\frac{m_u}{m_e} \sim \frac{1}{9} \left(9 \frac{m_t}{m_\tau} \right)^{\frac{1}{3}} \,, \tag{6.66}$$

where we have first removed the $\frac{1}{3} \times \frac{1}{3}$ factor from the m_t/m_τ ratio, and then reintroduced it after having taken the third root. This was required by the fact that these normalization factors, accounting for the passage from free quarks to $SU(3)$ singlets, don't enter in the contraction of phase sub-spaces. Putting all the informations together, we conclude that the phase-space sub-volume of the charged particles of the first family, $\mathcal{V}(u,d,e)$, should be given by:

$$\mathcal{V}(u,d,e) = 9 \, \frac{m_u}{m_{\nu_\tau}} \sim \alpha^{-1/3}_{SU(2)} \, \left(\sqrt{\alpha_{SU(2)}} \right)^{-1/3} \times \alpha^{-1}_{SU(2)} \,. \tag{6.67}$$

The second and third power of this volume give finally $\mathcal{V}(c,s,\mu)$ and $\mathcal{V}(t,b,\tau)$. To summarize, the mass ratios are:

$$\frac{m_\mu}{m_{\nu_\tau}} \sim \alpha^{-2}_{SU(2)} \,; \tag{6.68}$$

$$\frac{m_s}{m_{\nu_\tau}} \sim \frac{1}{3} \, \left(\sqrt{\alpha_{SU(2)}} \right)^{-2/3} \, \alpha^{-2}_{SU(2)} \,; \tag{6.69}$$

$$\frac{m_c}{m_{\nu_\tau}} \sim \frac{1}{9} \, \alpha^{-2/3}_{SU(2)} \, \left(\sqrt{\alpha_{SU(2)}} \right)^{-2/3} \, \alpha^{-2}_{SU(2)} \,; \tag{6.70}$$

$$\frac{m_\tau}{m_{\nu_\tau}} \sim \alpha^{-3}_{SU(2)} \,; \tag{6.71}$$

$$\frac{m_b}{m_{\nu_\tau}} \sim \frac{1}{3} \, \left(\sqrt{\alpha_{SU(2)}} \right)^{-1} \, \alpha^{-3}_{SU(2)} \,; \tag{6.72}$$

$$\frac{m_t}{m_{\nu_\tau}} \sim \frac{1}{9} \, \alpha^{-1}_{SU(2)} \, \left(\sqrt{\alpha_{SU(2)}} \right)^{-1} \, \alpha^{-3}_{SU(2)} \,. \tag{6.73}$$

These relations are completed by:

$$m_{\nu_\tau} \sim \frac{1}{2}\,\mathcal{T}_{(\mathrm{string})}^{-\frac{1}{2}} \times \left[\alpha_{SU(2)}^{-1}\right]^3 . \tag{6.74}$$

$$m_{\nu_\mu} \sim \frac{1}{2}\,\mathcal{T}_{(\mathrm{string})}^{-\frac{1}{2}} \times \left[\alpha_{SU(2)}^{-1}\right]^2 . \tag{6.75}$$

$$m_{\nu_e} \sim \frac{1}{2}\,\mathcal{T}_{(\mathrm{string})}^{-\frac{1}{2}} \times \alpha_{SU(2)}^{-1} . \tag{6.76}$$

As discussed in section 6.2.2, we expect to be able to normalize not the single electron's and neutrino mass, but the $SU(2)_{\mathrm{w.i.}}$-neutral combination (e, ν_e). Furthermore, what we should be able to normalize is not the pure electron's mass but that of the electrically neutral "compound" $(e, \bar{e})$. In the first case, we can say that $m_{(e,\nu_e)} \sim m_e + m_{\nu_e} \sim m_e$. In the second case, however, similarly to what happens with $SU(3)$ and the quarks, we expect $m_{(e,\bar{e})} \sim 2m_e$. Of course, analogous considerations apply to all families, and to quarks as well, because they are all charged also under $SU(2)_{\mathrm{w.i.}}$ and $U(1)_{\mathrm{e.m.}}$. If an almost logarithmic sequence of masses in passing from ups and downs of $SU(2)$ doublets allows in general to neglect the correction due to the lighter particle of the pair, what we cannot neglect is the factor 2 due to the fact that, as it was for the case of the non-perturbative mean scale, section 6.1.2, we are calculating the mass of a particle-antiparticle pair. As a consequence, we expect that with the formulae obtained in this section what we get is twice the mass of any state.

The values we obtain in this way are just the "bare" values of the mass ratios, the first step in the approximation, which must be improved by "actual time" corrections, in order to account for finer details of the phase spaces. We didn't consider yet the quark masses of the first family. As it appears from our discussion, the up quark seems to be heavier than the down quark, as it is reasonable to expect by analogy with the other families. However, this is wrong, as is also clearly indicated by the experimental observations. In the next sections we will pass to the explicit evaluation of all the mass values. We will there discuss also the corrections to the bare expressions, required by an improved description of the details of the string configuration. This is particularly necessary in order to discuss the masses of the second family, strongly affected by the "stable" mass scale of the Universe, the mean scale discussed in section 6.1.1, and the quarks of the first family. As we will see in section 7.1.2, what happens in this case is that consistency of the vacuum implies an exchange in the role of the up and down quark. The mass relations are shown in the table 10.

6.2.4 The $SU(2) \equiv SU(2)_{\Delta m}$ coupling

In order to compute masses, what it remains is to know the beta-function of the broken $SU(2)$ group which constitutes the basic ingredient of mass ratios. In principle, this is not the $SU(2)_{\mathrm{w.i.}}$ of weak interactions, which acts only on the "left-moving" part of the particles. We indicate it as "$SU(2)_{\Delta m}$", to distinguish it from the group of the weak interactions. In order to determine the $SU(2)_{\Delta m}$ beta-function, we cannot proceed as in the traditional approach, through a (perturbative) analysis of the spectrum and the corrections to the gauge coupling.

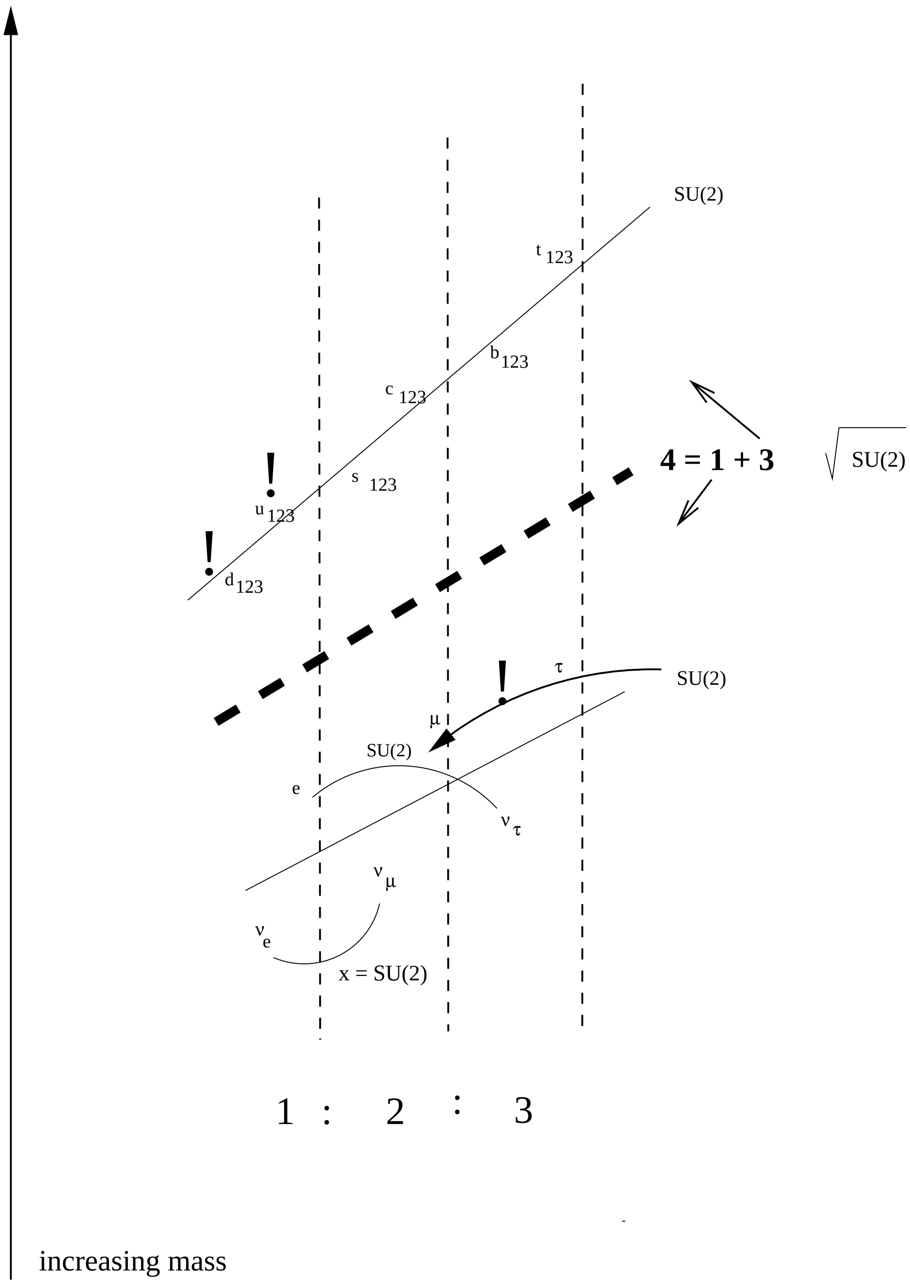


Figure 10: The diagram of elementary masses. Notice that up and down quarks are flipped. For leptons, the $SU(2) = SU(2)_{\Delta m}$ coupling factor separates the heaviest neutrino, ν_τ, from the lightest charged particle, the electron. This is due to the rearrangement of the $\mathbf{1}$ in the $\mathbf{4} = \mathbf{1} \oplus \mathbf{3}$, so that the three neutral particles are also the lightest ones, as required from entropy considerations.

Any perturbative computation, therefore performed in a logarithmic representation, does not simply account for the "logarithm of" the true beta-function: in any explicit string realization, part of the spectrum is perturbative and weakly coupled, but there is also a part which is "hidden", either because entirely non-perturbative, or because at least part of its interactions are non-perturbative. Therefore, it makes no sense to count the states of the spectrum, compute the beta-function as is usual in field theory, and then trade this quantity for the logarithm of the real beta-function.

In order to get this quantity, we proceed in another way. We have discussed how the effective coupling should account for the decay/interaction probability, and this should in turn be related to a volume occupied in the phase space. We expect that this is therefore valid also for a symmetry group: the strength of the corresponding interactions should be related to the volume occupied by the group in the phase space. For the computation of the beta-function a passage to the logarithmic picture is dangerous: we don't know what are all the states, perturbative and non-perturbative, of the spectrum, and we don't know how do they exactly appear in the logarithmic picture [47]. The analysis of the spectrum and the interactions carried out in section 3 was based on a comparison of several, dual "logarithmic pictures", or perturbative constructions, no one of them accounting for the full content of the theory. The gauge charges appeared in a rather different way on dual constructions. As long as it is a matter of counting the degrees of freedom, and listing their transformation properties, gathering together information obtained by "patching" dual pictures proved to be sufficient. However, in order to derive the strength of the couplings, the question is: how do we "patch beta-functions"? On the other hand, we have seen that, with a good degree of approximation, we can analyze the pattern of projections leading to the minimal entropy vacuum through the counting of Z_2 projections in orbifold representations. Through this procedure, non-perturbative sectors are automatically taken into account, being generated and/or projected out by operations that we can easily trace in some picture. In order to be sure to not forget, in the counting of the beta function, some states or misunderstand their role, we will therefore derive the volume occupied by the broken $SU(2)_{\Delta m}$ by counting the volume reductions produced by the various projections we have applied in order to reach the minimal entropy configuration.

The analysis of section 3 tells us that, within the conformal theory, we have at disposal 7 internal and 2 extended transverse coordinates for the projections. In total, we can apply 7 + 2 = 9 projections [48]. We must however consider that, for what concerns our problem, two of them don't reduce the volume of the phase space: with a first projection, supersymmetry is reduced from $\mathcal{N}_4 = 8$ to $\mathcal{N}_4 = 4$. With a second projection, supersymmetry is further reduced to $\mathcal{N}_4 = 2$, and new sectors of the spectrum are generated. However, in our specific case also at the $\mathcal{N}_4 = 2$ level the gauge beta functions vanish. This is not a general property of any $\mathcal{N}_4 = 2$ vacuum, but it is precisely what happens in the case of a configuration obtained with non-freely acting projections, even when coupled with freely acting, rank-

[47] Indeed, we will discuss in section 6.2.10 how in this picture the spectrum of the minimal entropy configuration appears to be supersymmetric.

[48] Said differently: we have room for shifts along 9 coordinates.

reducing shifts, as is our case [49]. Indeed, $\mathcal{N}_4 = 2$ is the step at which matter is generated in its full content, with the maximal amount of twisted sectors. For what concerns the gauge interactions, it is therefore the point of largest symmetry group. It is only through a further projection that, owing to the reduction to $\mathcal{N}_4 = 0$ [50], the gauge beta-function do not vanish anymore. Starting from this point, any reduction of the spectrum of matter states results in a corresponding reduction of the volume of the symmetry group. By counting the projections with the exclusion of the first two, we obtain that the last step, the one corresponding to the breaking to the minimal symmetry, corresponds, in the logarithmic picture, to a reduction of the volume of the phase space, with respect to the $\mathcal{N}_4 = 2$ configuration, by a factor $2 \times 7 = 14$. Through these projections the full initial symmetry group has *effectively* been reduced into a product of 14 equivalent factors, which correspond to $SU(2)$ enhancements of $U(1)$ symmetries. In other words, we can consider that the full symmetry of the phase space is a group G such that $G \supset U(2)^{\otimes 7}$. Each $U(2)$ factor rotates a subspace of dimension 2. Namely, on the tangent space (logarithmic representation) the "fundamental" representation is a direct sum of **2**:

$$\underbrace{\mathbf{2} \oplus \mathbf{2} \oplus \ldots \oplus \mathbf{2}}_{7}, \tag{6.77}$$

or, after the last step, in which also the $U(2)$ symmetry is broken and we remain with $U(1)$, a sum of 14 $U(1)$ representations:

$$\underbrace{\mathbf{1} \oplus \mathbf{1} \oplus \ldots \oplus \mathbf{1}}_{14}\,. \tag{6.78}$$

The "beta-function coefficient" (or better "exponent") of $SU(2)$ is then $\frac{1}{14}$ of the full exponent, which is fixed as follows. For $\mathcal{N}_4 = 2$ the beta function must vanish: no renormalization at all. The range of values is therefore $[0 - 1/2]$. According to expression 6.55, the beta-function exponent of $SU(2)_{(\Delta m)}$ is:

$$\beta_{SU(2)} = \frac{1}{14} \times \frac{1}{2} = \frac{1}{28}\,. \tag{6.79}$$

The coupling of $SU(2)_{(\Delta m)}$ is therefore:

$$\alpha_{SU(2)} = \mathcal{T}^{-\frac{1}{28}}\,. \tag{6.80}$$

Using the value of the age of the Universe given in appendix A, we obtain that, at the present day, $\alpha^{-1}_{SU(2)} \sim 147$. If more precisely we use the age of the Universe suggested by the agreement with neutron's mass, eq. 6.19 (i.e. $\sim 5,038816199 \times 10^{60} \mathrm{M}_{\mathrm{P}}^{-1}$, see Appendix A), we obtain:

$$\alpha^{-1}_{SU(2)} \sim 147,2 \; (147,211014)\,. \tag{6.81}$$

Being obtained through a counting of the orbifold projections, therefore not exactly at the real point corresponding to the minimal entropy configuration, 6.80 should constitute only

[49] See the discussion of section 3.2 and Ref. [11].

[50] We have seen that, although in some representation this configuration may appear as perturbatively supersymmetric with $\mathcal{N}_4 = 1$, supersymmetry is indeed broken.

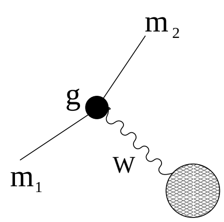
an approximation of the real value of the beta function. However, once again we expect the relative correction to $\beta^{-1} = 28$ to be small, of the order of the relative magnitude of an inverse root of the age of the universe, as compared to this integer value:

$$\beta^{-1} \approx 28 + \mathcal{O}\left(\alpha^{-1}\mathcal{T}^{-1/p_\beta}\right) , \tag{6.82}$$

which should reflect in a similar correction also for the coupling α ($\alpha^{-1} \to \alpha^{-1}(1+\mathcal{O}(1/\mathcal{T}^{1/p_\beta}))$).

6.2.5 The $U(1)_\gamma$ coupling

In order to obtain the coupling of $U(1)_\gamma$, the electromagnetic group, we don't need to determine the absolute fraction of a group factor within the full symmetry group: we can determine the ratio of the $U(1)_\gamma$ and $SU(2)$ phase spaces, or equivalently the ratio of the two exponents, by counting the charged matter states, and subtracting the number of gauge bosons. We can justify this if we consider that the latter contribute somehow "in opposite way" to the matter-to-matter scattering probability amplitudes. Consider a diagram corresponding to a matter-to-matter transition:



$$g_{\text{eff}} \cong \frac{g}{M_W^2} \tag{6.83}$$

For what concerns the initial and final matter states, we have that the larger the mass ratio between initial and final state, the larger is the decay amplitude. The boson mass appears instead at the denominator in the expression of the effective coupling, and suppresses the process.

A better way to see this is to consider that, as we will discuss in section 6.2.10, in the logarithmic picture the minimal entropy vacuum appears as effectively supersymmetric, with $\mathcal{N}_4 = 2$ extended supersymmetry. As seen from the logarithmic picture, the beta-function exponent is a $\mathcal{N}_4 = 2$ beta function coefficient. In this case $b = T(R) - C(G)$. An equal number of matter states and gauge bosons, transforming in the same representation, corresponds to an effective $\mathcal{N}_4 = 4$ restoration, a situation of non-renormalization, with vanishing beta-function exponent [51]. The phase space coefficient of $U(1)_\gamma$ is proportional to:

[51] In the philosophy of this analysis, when the number of bosons matches the number of matter states, this also means that the representation of the symmetry group is the same: a differentiation in the symmetries of gauge and matter states would correspond to a less entropic phase produced by some operation, even in the case this is not explicitly related to broken gauge states. This is for instance the case of the separations into different planes introduced by orbifold projections, where these mechanisms appear in a "frozen phase".

3(families) $\times$ 2($SU(2)$doublets) $\times$ ($\mathbf{1}+\mathbf{3}$)(leptons + quarks) $\times$ 2(left + right chirality) [= 48] $-$ 1(gauge boson) = 47. Notice that, in the counting, we have considered that *all* the matter states are charged under $U(1)_\gamma$. Three states, the three neutrinos, are however uncharged. However, the electromagnetic charge is simply "shifted" from the central value $(\frac{1}{2}, -\frac{1}{2})$, but the traceless condition is preserved. As a result, the charge is only "rearranged" among the states: some states result more charged, some less. In total, the strength of the renormalization is the same as with a traceless $U(1)$ with a charge equally distributed among all the states. This is true in first approximation: from a field theory point of view, this would be strictly true if all the masses of the matter states were the same, i.e. vanishing. Otherwise, the diagrams corresponding to the contribution of different particles have different amplitudes. From the point of view of this work, the corrections to this first order approximation are kept to a "minimal" degree by the requirement of entropy minimization, which implies a minimization of the total strength of the electro-magnetic interaction.

The beta-function coefficient of $SU(2)$ is proportional to 48 (the same effective number of states as for $U(1)_\gamma$) minus 3 (the number of gauge bosons), i.e. 45, where the coefficient of proportionality is the same as for $U(1)_\gamma$. The ratio of the two coefficients is therefore:

$$\frac{\beta_{U(1)}}{\beta_{SU(2)}} = \frac{47}{45}\,. \tag{6.84}$$

Using 6.80 and 6.84, and the scale $\mu = \mathcal{T} \sim 5,038816199 \times 10^{60} \mathrm{M}_{\mathrm{P}}^{-1}$, the present age of the Universe 6.19, adjusted on the neutron mass, we get:

$$\alpha_\gamma^{-1} \sim 183,777867\,. \tag{6.85}$$

The has to be considered as a "bare" value of the coupling, not an effective coupling in the field theory sense. We will discuss in sections 6.2.9 and 7.4 how this value should be "run back" to obtain the effective coupling to be compared with the value experimentally measured at a certain scale.

6.2.6 The $SU(2)_{\mathrm{w.i.}}$ coupling

Determing the coupling of the $SU(2)$ of the weak interactions is even more problematic than determining α_γ. The point is that for us this symmetry is not spontaneously broken in the classical sense, and we cannot compute the beta-function coefficient in an effective theory with unbroken gauge symmetry. In the usual field theory approach, the $SU(2)$ acting on just one of the two helicities transforms only half of the matter degrees of freedom, and therefore, if we neglect the contribution of the gauge bosons, its beta-function coefficient turns out to be one half of that of a "full" gauge group, namely, with a vectorial coupling to the matter currents. This is however true as long as the matter states are massless (on the other hand, once they acquire a mass, the gauge symmetry is broken). Massive states consist of both left and right degrees of freedom. From the point of view of the volume occupied in the phase space, although interacting with just their left-handed part, massive matter degrees of freedom count as much as left + right chiral states. The volume

occupied by $SU(2)_{\rm w.i.}$ is therefore something "intermediate" between the situation of pure chiral gauge symmetry acting on massless states, therefore on half the space of the degrees of freedom, and a full vectorial interaction. We don't know a rigorous way of counting the volume of this interaction in the phase space. If we want just to give a rough estimate, we can approximately consider that our coupling lies somehow "in between" the two situations: since in first approximation the matter states acquire a mass through a shift that reduces by half the logarithmic volume of the space (resulting therefore in a square-root scaling law), we can expect that the logarithmic volume occupied by $SU(2)_{\rm w.i.}$ is the mean value between the one of the vectorial interaction (acting therefore on the same number of degrees of freedom as the massive matter states), and the one of the pure chiral interaction, viewed as acting on massless states:

$$\beta_{SU(2)_{\rm w.i.}} \approx \frac{1}{2}\left(1 + \frac{1}{2}\right) \times \frac{1}{28}\,. \tag{6.86}$$

The present-day value of the inverse of the $SU(2)_{\rm w.i.}$ coupling should therefore be:

$$\alpha_w^{-1} \approx \mathcal{T}_0^{-\left(\beta_{SU(2)_{\rm w.i.}}\right)} \sim 42,26\,, \tag{6.87}$$

where we have used the estimate of the age of the universe 6.19. The value 6.87 is roughly a factor 4,4 smaller than the inverse electromagnetic coupling, given in 6.85. Also this number has to be considered a "bare" value, to be corrected in the way we will discuss in section 6.2.9.

6.2.7 *The strong coupling*

In our framework, the $SU(3)$ colour symmetry is always broken, and in principle there is no phase in which the strong interactions can be treated as gauge field interactions at the same time as the electromagnetic ones. In particular, there is no (under-Planckian) phase in which the strongly coupled sector comes down to a "weak" coupling, which merges with the other couplings of the theory to build up a unified model with a unique coupling, taking up the running up to the Planck scale. For us, the strongly coupled sector is strongly coupled at any sub-Planckian, i.e. field theory, scale. The coupling α_s will always be larger than one:

$$\alpha_s \sim \mathcal{T}^{-\beta_s}\,, \qquad \beta_s < 0\,. \tag{6.88}$$

Indeed, the representation in terms of an $SU(3)$ gauge symmetry is something that belongs more to an effective field theory realization than to the non-perturbative string configuration we are considering. Namely, in our case we just know that, as soon as the space is sufficiently curved (i.e. entropy sufficiently reduced), we have the splitting into a weakly and a strongly coupled sector, mutually non-perturbative with respect to each other.

In order to derive the exponent β_s, we must proceed as in section 6.2.4, by computing the amount of symmetry reduction, this time however in the "S-dual" representation. As we discussed in section 3, when seen from the point of view of the full space, this duality is indeed a T-duality . This is basically the reason why the coupling increases as the temperature of the Universe decreases (or equivalently its volume increases). We expect therefore that, when seen from the point of view of a dual picture, the coupling arises in a vacuum which

underwent the same amount of symmetry reduction as in the case of the dual $SU(2)$ case of section 6.2.4. However, the space-time coordinates feel a "contraction" which is T-dual to the one experienced in the picture of the electro-weak interactions. Therefore, when referred to the time scale of the "electroweak picture", the "beta-function" exponent, the coefficient β_s, should be 1/4 of its analogous given in 6.79. Of course, as seen from the electroweak picture, the sign is also the opposite (an inversion in the exponential picture reflects in a change of sign of the logarithm). We expect therefore:

$$\beta_s = -\frac{1}{4} \times \frac{1}{28}\,. \tag{6.89}$$

In other words, the strong coupling in itself should run as:

$$\alpha_s \sim (\mathcal{T}_{\text{dual}})^{\frac{1}{28}}\,, \tag{6.90}$$

but the time scale $\mathcal{T}_{\text{dual}}$ is related to $\mathcal{T}$ by an inversion *times* a rescaling. As the value 6.80 can be seen as the "on-shell" value at the matter scale $1/\sqrt{\mathcal{T}}$, logarithmically rescaled by a factor 1/2 with respect to the un-projected time scale $\mathcal{T}$, the scale $\mathcal{T}_{\text{dual}}$ feels an inverse logarithmic rescaling, $(1/2)^{-1}$. In total, as compared to the square-root scale, it has a logarithmic rescaling by a factor 4. In order to refer the value of the strong coupling to the square-root scale of the electroweak picture, we must therefore take its fourth root. The present-day "bare" value of the strong coupling is therefore [52]:

$$\alpha_s|_{\text{today}} \sim \left[(\mathcal{T}_0)^{\frac{1}{4}}\right]^{\frac{1}{28}} = \mathcal{T}_0^{-\beta_s} \sim 3,48\,. \tag{6.91}$$

As in the case of α_γ and α_w, in order to be compared with the coupling currently inserted in scattering amplitudes also this one has to be "run back" in the way we will discuss in section 6.2.9 .

6.2.8 The "unification" of couplings

As one cas see, in our framework the fundamental scaling of couplings as a power of the age of the Universe does not involve the "field theory gauge strength" g, the strength of the gauge covariant derivative, which couples the gauge field to the matter kinetic term, but the quantity $\alpha = g^2/4\pi$. On the other hand, it is α the physical coupling entering in any expansion, always given as a series of powers in $g^2/4\pi$. As a consequence, for us the quantities which go to 1 and unify, in the specific case at the Planck scale, are not the three couplings g_s, g_1, g_2 introduced through a gauge mechanism, but the effective strengths entering in scattering and decay amplitudes, namely the three couplings α_s, α_w and α_γ:

$$\alpha_i = \mathcal{T}^{-\beta_i} \implies \lim_{\mathcal{T}\to 1} \alpha_i = 1\,. \tag{6.92}$$

This in particular means that, at a certain scale, close to but below the Planck scale, the couplings g of the electro-weak interactions will become "strong": $g > 1$. This however

[52] Also in this case we don't need a high precision in the estimate of the age of the universe, whose value appears here rather suppressed.

doesn't mean that the corresponding interactions are going out of the weak coupling regime. In our framework, strictly speaking there are no gauge interactions, the gauge representation being only a useful approximation, and the gauge connection g doesn't have a particular physical meaning, besides being a useful tool in order to arrive to α. On the other hand, even at the "m_Z" mass scale, α_s is in our case larger than one. What is then the meaning of a $\sim 0,2$ value for this coupling, as predicted by the usual $SU(3)$ colour analysis, and "confirmed" by experiments? In our case, such a value would mean that the strong interactions are not strong at all, but well perturbative, as are the electromagnetic and the weak one. In the next section, we will discuss how the values we have obtained for the three couplings do compare with the parameters of an effective action, and therefore with the data one finds in the literature.

6.2.9 The effective couplings: part 1

The couplings α_γ, $\alpha_{\text{w.i.}}$ and α_s we have derived in section 6.2.4 and 6.2.5 and 6.2.7 run with time, and therefore with an energy scale, but not in the usual sense of the renormalization group. Namely, they are the *fixed* couplings at a specific age of the Universe. The "electron mass scale", or the "Z-boson scale", here would mean a different age, and size, of the Universe. The usual running according to the equations of a renormalization group refers instead to the "effective" rescaling in a Universe whose fundamental parameters remain fixed. This is due to the fact that space-time, the space in which the effective action is framed, is normally assumed to be of infinite extension. The infinities which are in this way produced in the effective parameters must be regularized according to certain rules; in practice, by keeping as reference points certain "on shell", "physical" values. If we want to compare our results with the parameters of such an effective action, we must take into account this mismatch in the interpretation of space-time. Namely, we must correct for a finite extension of the Universe. This is necessary in order to compare with the experimental data as they are quoted in the literature. Any experimental value is derived in fact through comparison of a certain "scattering amplitude" with an effective formula, expressed in terms of couplings, masses, momenta etc... For instance, the value of the fine structure constant is obtained from a process taking place at the electron's scale. The amplitude is computed by integrating over the momentum/space-time coordinates. In the traditional field theory approach, this is done in an infinitely extended space-time. In our case, instead, the volume of space-time is finite; the fraction of the moduli space occupied by the process is therefore relatively higher as compared to the case of infinite volume, and the probability of the transition too. We understand therefore how it is possible that, in our framework, the same amplitude is obtained with a smaller electromagnetic coupling as in the usual approach, where the volume of space-time is infinite.

The correction due to the finiteness of space-time involves not only couplings, such as those of the electro-magnetic and strong interactions, but also, as a consequence, the masses. Indeed, when we say that an elementary particle has a certain mass, it is always intended that this is the "on shell" mass of the particle "at rest", and considered as an asymptotic, free state. For instance, in the case of the electron this means that it is considered "at the

electron's scale", and in a phase in which it can be assumed to be decoupled from the other particles, i.e. in a weak coupling regime. From our point of view, this means in a decompactification phase. However, in our scenario the true, physical decompactification occurs only at the infinite future. The decompactification implied in such arguments is therefore an artifact: one takes a limit of weak coupling of some string coordinate, a linearization that corresponds to working on the tangent space, while curing the so produced infinities/zeros and trivializations by imposing a regularization procedure, a renormalization prescription. In our case, we keep on imposing that the neutron's mass is the one given as in 6.17. This means, trating the neutron's mass as an already *renormalized* value, and considering the relation 6.17 as an "on shell prescription" which we use in order to fix the regularization. The finiteness of the space volume can then be taken into account by considering any mass and coupling in a renormalization group analysis in which we use a finite-volume regularization scheme. The physical masses and couplings become therefore scale dependent, i.e. this time not only they depend on the age of the universe, by prescription fixed by the value of the neutron's mass, but also on the scale at which the process they correspond to takes place. Couplings and masses have therefore a ground time-dependence, as a power of the age of the universe, as a consequence of the time-dependence of the renormalization prescription, and a milder, logarithmic dependence on the cut-off scale of the renormalization.

The electromagnetic and weak couplings

To start with, in this section we consider the correction to the weak "gauge" couplings [53]. The correction to the $SU(2)_{\Delta m}$ couplings, and to the masses, will be considered in a further section 7.3. In the representation at infinite volume, the effective gauge couplings are corrected according to:

$$\alpha_i^{-1} \approx \alpha_i^{-1}|_0 + b_i \ln \mu / \mu_0 \,, \tag{6.93}$$

where b_i are appropriate beta-function coefficients, and μ is the scale of the process of interest (this can be the electron mass in the case of the fine structure constant). In principle, in this vacuum there is no Higgs field, however it is not clear what should be the best linearized representation of the physical configuration: is it non-supersymmetric, as in the "real" picture, or does the process of linearization lift down some supersymmetry? How does one correctly *approximate* the physical situation, which in itself escapes the rules of field theory, with a field theory in which to consistently perform computations? It could be that the introduction of a Higgs field *mimics* with a certain accuracy the effect of masses, in the practical purpose of computing the running of effective parameters in the neighbourhood of a certain scale of the Universe. There is therefore an uncertainty in the definition of the running. However, as a consequence of 6.92, in first approximation we can assume that, in the effective representation of the physical configuration, couplings run logarithmically with an effective beta-function such that, starting from their "bare" value at the actual $\mathcal{T}^{-1/2}$ scale, they meet at zero at the Planck scale:

$$\alpha_i^{-1} \approx \alpha_i^{-1}|_0 + b_i^{(\mathrm{eff.})} \ln \mu / \mu_0 \,, \tag{6.94}$$

[53] The strong coupling α_s requires a separate discussion.

with $b_i^{(\text{eff.})}$ such that:

$$b_i^{(\text{eff.})} \ln \mu_0 \;=\; \alpha_i^{-1}|_0 \,. \tag{6.95}$$

The scale μ_0 is fixed to be the end scale of our symmetry reduction process, the one corresponding to the lowest level attained with the projections, the square root scale:

$$\mu_0 \;=\; \frac{1}{2}\mathcal{T}^{-\frac{1}{2}}\,, \tag{6.96}$$

where $\mathcal{T}$ is the age of the universe as fixed by the neutron's formula 6.17. The choice of the square root scale 6.96 as the starting scale is dictated by the fact that this is the fundamental scale of matter states, and their interactions. Matter consists of spinors and their compounds, and a spinor feels a square-root space, in that twice a spinor rotation corresponds to a true vectorial space rotation. From a technical point of view, the square root scale is the one produced by the shift giving rise to masses for the matter states in this string scenario. As we will discuss in section 7.3, the exact normalization of the end scale for elementary states is 1/2 of 6.96.

Let's consider the electromagnetic coupling. The value of α_γ given in section 6.2.5 must be considered as a bare value at the scale μ_0. The fine structure constant, which for us is not really a constant, it is just the present-day value of this coupling. will correspond to the value of α_γ run from 6.85 at the scale 6.96 to a scale μ_γ, typical of some process related to the electric charge. As it is experimentally given, this quantity refers to the scale of the electron at rest. This is on the other hand the original scale at which historically the electric charge has been referred to. Although in general modern experiments are not performed at the electron's scale, through renormalization techniques their measurements are anyway always reduced to the electron's scale. From the point of view of our theoretical framework, this is the scale at which the "charged world" starts. Below this scale, there are the un-charged particles, and, from a classical point of view, the electric charge effectively ceases to exist. Once recalculated on the electron's mass scale, 6.85 gets corrected to:

$$\alpha_\gamma^{-1} \;:\; \alpha_\gamma^{-1}|_{\mu_0} \;=\; 183,78 \;\to\; \alpha_\gamma^{(0)\,-1}|_{m_e} \;\approx\; 132,85\,, \tag{6.97}$$

where we used the value 7.36 for the electron's mass. The result 6.97 is definitely closer to the experimental value, nevertheless still quite not right, being out for an amount higher than the error in our approximations. The reason is that the value 6.97 has been calculated by assuming a perfect logarithmic running, without taking into account for an important modification in the volume of the phase space of the charged matter particles around the electron and up quark mass scale, something we will do in section 7.1.2. We postpone therefore a detailed evaluation of the fine structure constant to section 7.4.

For what matters the weak coupling, the contact with experiment is made through the Fermi coupling constant G_F, basically the weak coupling divided by the W-boson mass squared. Any discussion about this must therefore be postponed after we have obtained this mass. However, the W mass too, in order to be calculated, requires a first order estimation of the weak coupling. Indeed, proceeding as in 6.97, we can see that also this coupling undergoes relative corrections of the right magnitude. We will come back to this coupling in section 7.7.

The strong coupling

In the case of the strong coupling, things are, for obvious reasons, more involved, being more model-dependent also the theoretical framework in which its effective experimental value is obtained. A possible "contact with the experiment" is the value α_s at the scale of some typical quark process, for instance the Z-boson mass in a $e^+e^- \to 4\mathrm{J}$ event: $\alpha_s(M_Z) = 0,119$ [58]. As it is usually given , g_s runs logarithmically with the scale. It seems therefore impossible to think that the "on shell" value 6.91 can be effectively corrected to a current value lower than 1 at around 100 GeV. However, not necessarily α_s must admit an effective representation in terms of a logarithmic running *at the same time*, i.e. *in the same picture* as the electromagnetic and weak couplings. Namely, although strongly and weakly coupled sectors are usually described in an effective action that accounts for all of them at the same time, attributing a logarithmic running to all of them in a unified picture, there are good reasons to believe that, especially for low energies, in the case of the effective α_s the logarithmic behaviour is only a first approximation. Indeed, electro-weak and strong coupling are mutually non-perturbative with respect to each other, and, although we don't know what the correct resummed running of α_s should be, and we can only make some speculation, we may expect that its logarithmic behaviour is only the first order approximation of a running that, in the representations in which the electro-weak couplings are linearized, is exponential. If we suppose that the amount of change computed in a certain scale interval should be seen as the first step of an exponential correction, namely, if we suppose that it increases/decreases by a factor ~ 15 for each $\Delta\mu \approx 10^{12\sim 13}\,\mathrm{M_P}$, then it is not impossible that, in passing from the scale $\mu_0 \sim 10^{-30}\,\mathrm{M_P}$ to $\sim 10^{-17}\,\mathrm{M_P}$ (~ 100 GeV) the value of the strong coupling passes from 6.91 to $\sim 0,2$. It appears therefore that an effective value of α_s lower than 1, as it is usually obtained, is not a signal of weakness of the interaction, but the result of working in a "fictitious", infinitely extended space-time.

6.2.10 Running in the "logarithmic picture"

In our analysis, we have used several times the mapping to an artificial logarithmic representation of space-time, in order to investigate properties related to a vacuum that, in its correct, "physical" representation, appears to be strongly coupled. This proves to be particularly useful in the investigation of some properties of leptons and quarks as free states. In the logarithmic picture, a coupling of order one becomes weak, and therefore strongly coupled elementary states become weakly coupled. When considered at early times, namely at a stage in which even the weak interactions are "strong", this representation allows to deal at the same time with leptons and quarks. This is not something so unfamiliar: although not stated in these terms, all the perturbative, either field theoretical or stringy, constructions of a spectrum such as that of the Standard Model of electro-weak interactions are based on the assumption of working in such a kind of representation.

What we did not yet discuss is how does the spectrum of the low-energy theory look in such a representation, as compared to the "exponential" picture, the physical one. In section 3.2.1 we discussed how supersymmetry breaking is related to the appearance of a non-vanishing curvature of space-time. This effect can therefore be viewed as "tuned" by a

"coordinate" of the theory, that remains twisted, and therefore frozen, at the Planck scale. Since the strength of the coupling of the theory is related to the volume of the "internal", i.e. non-extended, coordinates, the breaking of supersymmetry is a phenomenon related also to the appearance of strongly coupled sectors, and strongly coupled matter states. Roughly speaking, we could say that strong coupling and supersymmetry breaking go together and are tuned by the same parameter. This is however a "composite" parameter (such as a function of the product of coordinates).

We have seen that logarithmic pictures, and more in general, perturbative constructions, correspond to "decompactifications" of the theory. The decompactification can be either a true flat limit, or be just a singular, non-compact orbifold, which appears flat only locally and perturbatively.

In both cases, the theory doesn't appear higher dimensional, because this operation involves an "internal" coordinate, the coupling of the theory. Owing to the flattening, the "trivialization" of this parameter, from a perturbative point of view it appears as a limit of partial restoration of supersymmetry. However, strictly speaking this is not a smooth limit of the string space, which remains basically twisted. That's why we prefer to speak in terms of logarithmic mapping rather than of "limit". The amount of restored supersymmetries depends on the details of the way the "decompactification" is obtained. Indeed, a full bunch of over-Planckian states could in principle come down to a light mass: extended supersymmetries could appear, as well as the extra states lifted by the rank reductions. However, for the investigation of the properties of matter states, the logarithmic mapping of interest for us is the "minimal" one, such that just one parameter, the one responsible for the separation of weakly and strongly coupled sectors, is mapped to zero, while all the other internal coordinates remain "twisted". Depending on the "dual" representation we want to consider, the amount of supersymmetry we are going to recover in the logarithmic picture is therefore either $\mathcal{N}_4 = 1$ or $\mathcal{N}_4 = 2$. We already discussed that $\mathcal{N}_4 = 1$ is a fake, unstable configuration consisting of the projection onto just the perturbative part of the spectrum of a theory which, non-perturbatively, is non-supersymmetric. In the case we are interested in a correct understanding of the beta-functions, mapping to a $\mathcal{N}_4 = 2$ logarithmic representation is more appropriate than to a "fake" $\mathcal{N}_4 = 1$. This is what we have done in section 6.2.5, in order to understand the role played by matter states and gauge bosons in the evaluation of the $U(1)_\gamma$ beta-function as compared to the $SU(2)$ beta-function. In this representation, there is no "parity restoration" in the sense of the $SU(2)_{(\mathrm{R})}$ bosons coming to zero mass. The fact that $\mathcal{N}_4 = 2$ supersymmetry doesn't have a chiral matter spectrum (hypermultiplets include the conjugate states of fermions) simply means that we must expect a doubling of the matter states, due to the fact that both the left and right moving part of a matter state get paired to a conjugate. On the other hand, this is not a problem, because this picture is just a useful representation, in which we can understand certain properties, that must however be appropriately pulled back to the physical picture. In the computation of the beta-function coefficients we don't need to consider this doubling of degrees of freedom, because this is also related to an effective disappearance of one of the projections.

6.2.11 *Recovering the "Geometric Probability" tools.*

Our method of deriving masses and couplings through an analysis of the associated entropy of the phase space can be viewed as a "lift up" of the methods of deriving these quantities through a computation of the geometric probability of the interaction processes. The idea goes somehow back to the work of Armand Wyler [63], in which the value of the fine structure constant is given as a ratio of volumes, which can be interpreted as phase space volumes. Further developments have shown how, through an appropriate representation of the phase space of particles and interactions, it is possible to obtain couplings and masses which are extremely close to the experimental ones [64, 65, 66, 67]. However, these are given as pure numbers, no dependence on the fundamental scale of the Universe seeming to be implied. In order to understand this point, we must consider that, from the point of view of this work, all these computations are performed in *linearized* representations of the physical space. Even in the case space-time, and the associated phase space, is seen as the "tangent space" of an embedding curved space, this higher space is somehow just the "minimal" embedding space, a "first order" departure out of the base, the flat four dimensional space. All this to say that, from our point of view, these analyses are a kind of "logarithmic picture analyses". Let's consider what usually happens to the coupling. In our framework, it scales as a power of the age of the Universe; in a logarithmic representation, as a logarithm of. Nevertheless, up to a certain extent it is possible, and it does make sense, to reparametrize the scale dependence by approximating a power-law dependence with a logarithmic one, in such a way that:

$$\alpha^{-1} = \left(\frac{1}{\mathcal{T}^{p_\alpha}} \right)^{-1} \quad \leftrightarrow \quad \approx \frac{1}{\alpha_0} + \beta \log \mu \,, \qquad \alpha_0^{-1} \equiv \beta \log \mu_0 \,, \quad \mu \equiv \mathcal{T} \,, \tag{6.98}$$

for some value of α_0 and β. On the left hand side, we have the real running in the physical picture (the "cosmological picture", to speak), on the right hand side we have the usual running in a perturbative, effective field theory representation. We have discussed in a previous section (6.2.9) how and why this basically works. In sections 6.2.1–6.2.7 we have also mentioned the problem of computing the exponents to which the age of the Universe must be raised, alluding at the possibility of calculating their ratios in a logarithmic representation. Of course, in such an effective representation, the coefficient β is not the same number as the exponent p_α on the l.h.s. of 6.98. For instance, we have seen in 6.80 that the $U(1)$ exponent is $\approx 1/28$, which is quite far from the electro-weak beta-function coefficients of ordinary field theory. This because the r.h.s. of expression 6.98 is an effective reparametrization of the physical problem, not simply the logarithm of the l.h.s. From this point of view, the Wyler's formula for the electromagnetic coupling is the measure, in a linearized, logarithmic representation, of the volume corresponding to this coupling in units of the volume of a five-dimensional space. To be concrete, using the value 6.82 we obtain:

$$\frac{\left[\alpha^{-1} \approx (\mathcal{T}^5)^\beta \right]}{V(\mathcal{T}^5)} \quad \overset{\log}{\rightsquigarrow} \quad \simeq \frac{\beta \log(\mu/\mu_0)}{\log(\mu/\mu_0)} \,, \qquad \beta \approx \frac{1}{5} \times \frac{1}{28} = \frac{1}{140} \,, \tag{6.99}$$

where μ/μ_0 corresponds to $\mathcal{T}^5$ and $\beta \approx \frac{1}{140}$ is here our rough approximation of the fine structure constant, what in Wyler's formula indeed comes out closer to the experimental

value. This would probably happen also in the present case, after a more accurate inclusion of the corrections discussed in section 6.2.9. The fact that with this procedure one gets a number that corresponds to the electromagnetic coupling, which can be interpreted as a geometric probability in a five-dimensional embedding of the four-dimensional physics [67, 66] is then here no more than a matter of coincidence: the coupling evolves with time, and "measuring" the exponent 6.80 in a five-dimensional space is just a lucky choice, which works because, at present time, $\mathcal{T} = \mathcal{T}_0 \sim 5 \times 10^{60}\,\mathrm{M}_\mathrm{p}^{-1}$, indeed $\mathcal{T}_0^{-(1/28)} \sim \frac{1}{5} \times \frac{1}{28}$. However, all this acquires a deep meaning when the so derived couplings and masses are instead seen as ratios of volumes, normalized to a specific scale ([64, 65, 66]). If we consider ratios of couplings in a "logarithmic representation", it is easy to see that any scale dependence drops out. At the ground of the disappearance of any scale dependence is the fact that all the quantities we deal with in field-theory effective representations are regularized quantities. Namely, one performs the analysis around a starting point, such as α_0 on the r.h.s. of expression 6.98, a value obtained after regularization of the infinities, "endemic" of a representation in an infinitely extended space-time. One deals then with constant, regularized values, and perturbations around the regularization point. The scale dependence appears only as a correction to a scale-independent bare value. Although corresponding to an artificial reparametrization at a certain fixed point of the cosmological evolution, such a representation of the physical space can anyway be very useful. In principle, with our approach one catches the full behaviour of masses and couplings. In particular, we get the running along the history of the Universe, something essential in order to understand astronomical experimental observations, or more in general the physics of much earlier times of the history [54]. In practice however, in order to perform fine computations limited to our present time, it may turn out convenient to map to an appropriate "linearized" representation, such as those considered in Refs. [65, 66, 64], in order to carry out a refined calculation of masses and couplings.

[54]See sections 9.2, 10.1, 10.2 and 10.3 for a discussion of these issues.

7 Current mass values of elementary particles

Now that we have at hand the value of the $SU(2)_{\Delta m}$ coupling, we can proceed to an explicit evaluation of the masses of the elementary particles, listed in table 10. Free elementary particles correspond to our conceptual classification more than to the real world. As we mentioned in section 6.2.2, asymptotic running mass formulae are naturally given for states which are neutral under the three elementary interactions, all strong at the Planck scale. At present time, leptons are weakly coupled, while quarks feel a coupling which is even stronger than it was at the Planck time. In order to obtain the mass of the elementary particles, intended as free states, we must therefore identify what are the "minimal composite states", i.e. the lightest singlets they can form. We will in the following proceed by discussing the situation case by case.

7.1 "Bare" mass values

We consider here the mass values corresponding to the elementary particles, as they can be computed using the mass formulae given in section 6.2.3. These can be considered the "bare" values, to be corrected in various ways, in order to account for the mismatch between the finite-volume and the usual infinite-volume approach, and for the fact that at any finite time completely free states are an ideal representation, but don't really exist. Mass scales can therefore be perturbed by the "stable" mass scale 6.17. In section 7.2 we will discuss these corrections, and how, in some cases, it is even more appropriate to consider these values themselves as "corrections" of a "bare" mass scale.

7.1.1 Neutrino masses

We start with the less interacting, and therefore lightest, particles. According to the considerations of sections 6.2.1 and 6.2.3, the lightest mass level must correspond to the lightest electrically neutral particle, the electron's neutrino. Using the value of the present-day age of the Universe derived from the neutron's mass, expression 6.19, we obtain the following value for the "square root scale" :

$$\frac{1}{\mathcal{T}^{1/2}} \approx 4,454877246 \times 10^{-31} \mathrm{M_P}\,. \tag{7.1}$$

Following 6.59 and 6.55, the first neutrino mass should be a $\alpha^{-1}_{SU(2)}$ factor above 1/2 this scale. Furthermore, as discussed in section 6.2.3 after the expression 6.76, this procedure, being related by a chain of symmetry reduction factors to the mass of an electrically neutral electron-positron pair, gives twice the mass of the neutrino, or, better, the $\nu\bar{\nu}$ mass. Using the value 6.81, we obtain therefore:

$$2\, m_{\nu_e} \approx 3,279 \times 10^{-29} \mathrm{M_P} \sim 4,0037 \times 10^{-10} \mathrm{GeV} = 0,40\, \mathrm{eV}\,. \tag{7.2}$$

After multiplication by a further $\alpha^{-1}_{SU(2)}$ factor, we obtain the second neutrino mass:

$$2\, m_{\nu_\mu} \approx 5,89 \times 10^{-8} \mathrm{GeV} = 58,9\, \mathrm{eV}\,. \tag{7.3}$$

Finally, multiplication by a further $\alpha^{-1}_{SU(2)}$ factor leads us to the tau neutrino:

$$2\, m_{\nu_\tau} \, \approx \, 8,677\, \text{KeV} \, . \tag{7.4}$$

These values agree with the experimental indications of possible neutrino oscillation effects at the electronvolt scale.

7.1.2 The charged particles of the first family

An $\alpha^{-1}_{SU(2)}$ factor above the mass of the tau-neutrino there is the electron's mass:

$$m_e \, \sim \, \alpha^{-1}_{SU(2)} \, \times \, m_{\nu_\tau} \, \sim \, 0,639\, \text{MeV} \, . \tag{7.5}$$

As discussed in section 6.2.3, this should be the mass of an electron-neutrino compound. However, as we have seen, neutrino masses are negligible in comparison to lepton masses, and with a good approximation the mass of such a compound coincides with the lepton's mass.

Continuing along the lines of section 6.2.3, from 6.67 we should be able to derive then the down and up quark masses, obtaining $m_d \sim 0,48\, \text{MeV}$ and $m_u \sim 0,87\, \text{MeV}$. However, this is not correct, and is contradicted by the experimental observations. The explanation has to do with the way the symmetry breaking is realized in our framework. At low energy, the $SU(2)_{\text{w.i.}}$ symmetry appears as a broken gauge symmetry, with the breaking tuned by a parameter of the order of a negative power of the age of the Universe. As we will see in section 7.6, the $SU(2)_{\text{w.i.}}$ gauge boson masses scale in such a way that $\mathcal{T} \to \infty$ is a limit of approximate restoration of the $SU(2)_{\text{w.i.}}$ symmetry. Moreover, remember that the weak force in itself is stronger than the electromagnetic force: $\alpha_w > \alpha_\gamma$ (it is called weak because for low transferred momenta, $p/M_W \ll 1$, effective scattering/decay amplitudes are suppressed by the boson mass: $\alpha^{\text{eff}}_w \approx \alpha_w / M_W$). Therefore the "hierarchy" of matter is prioritarily determined by the $SU(2)_{\text{w.i.}}$ charge, more than by the electric charge. As a consequence, the matter spectrum can be thought as made of two subspaces, the "up" and the "down" subspace, and the trace of the electric charge can be viewed as:

$$< Q_{e.m.} >= \sum_{\ell,\text{q}} < \text{up}|Q_{e.m.}|\text{up} > \, + \sum_{\ell,\text{q}} < \text{down}|Q_{e.m.}|\text{down} > , \tag{7.6}$$

where $\sum_{\ell,\text{q}}$ indicates the sum over leptons and quarks. As we discussed in section 3.3, minimization of entropy requires to choose a particular distribution of the electric charge among the $SU(2)_{\text{w.i.}}$ singlets (up + down), so that we can have an electrically neutral particle. The condition of approximate restoration of the $SU(2)_{\text{w.i.}}$ symmetry, and the dominance of the weak force with respect to the electromagnetic one, require that the two terms of the r.h.s. of 7.6 give an equal contribution to the total mean value of the electric charge. Otherwise, this would explicitly break the $SU(2)_{\text{w.i.}}$ invariance. This imposes that the trace of the electric charge has to vanish separately on the "up" and "down" multiplets. In practice, both of them must vanish.

For the validity of this argument it is essential that the weak force ends up by dominating the more and more over the electric one, and that the symmetry is restored at infinitely

extended space-time; therefore, the full space must be essentially thought as separated in two $SU(2)_{\text{w.i.}}$ eigenspaces. Compatibility of the theory at any finite time with the situation at the limit tells us that:

$$\operatorname{tr}(\nu, d) \,=\, 0\,. \tag{7.7}$$

Since the ν charge vanishes, we have that:

$$\operatorname{tr}(d) \,=\, 0\,. \tag{7.8}$$

This is only possible if, for one family, the roles of the up and down quarks, for what matters the electric charge, are exchanged, so that we have $\operatorname{tr}(d) = 3 \times \left(\frac{2}{3} - \frac{1}{3} - \frac{1}{3}\right) = 0$. Correspondingly, the trace of the "ups" is also vanishing:

$$\operatorname{tr}(\mathrm{e}, \mu, \tau, \mathrm{u}) \,=\, -1 \,-1 \,-1 \,+\, 3 \times \left(-\frac{1}{3} + \frac{2}{3} + \frac{2}{3}\right) \,=\, 0\,. \tag{7.9}$$

Therefore, in one of the three quark families the role of up and down is interchanged: the quark with electric charge +2/3 is indeed the "down", while the one with charge -1/3 is the "up". In the ordinary field theory approach, this argument does not apply because the symmetry remains broken also at infinitely extended space-time [55].

Simple entropy considerations allow us to identify in which family the flip occurs. Let's consider the $SU(3)_c$-singlet made out of the three quarks, one per each family, with higher electric charge, and the one made in a similar way out of the three quarks with the lower electric charge. Clearly, the first one is the most interacting singlet we can form by picking one quark from each family, and conversely the other one is the less interacting one we can form. The first must therefore also be the most massive out of all the possible $SU(3)$-singlets formed by one quark per each family, while the second one must be the lightest. The only possibility we have to achieve this condition is when the flip between charge +2/3 and -1/3 quarks occurs in the lightest family, i.e., for the quarks we usually call the up quark and the down quark. Therefore, approximately the value of the mass of the up quark is the one we computed for the lightest "down" quark states, and conversely the mass of the down quark is the one we assigned to the lightest "up". However, now the lightest quark has a higher electric charge. Namely, from charge $|Q| = \frac{1}{3}$ we pass to $|Q| = \frac{2}{3}$. This transformation is *not* a rotation of the group $SU(2)_{\text{w.i.}}$, but a pure electromagnetic charge shift. Therefore, here it does not matter that the former down had negative charge, so that the charge *difference* is $\Delta Q = \frac{2}{3} - \left(-\frac{1}{3}\right) = 1$: a charge conjugation is a symmetry for what matters the occupation in the phase space, or equivalently the mass. What counts is the pure increase in the absolute value of the charge, which implies an increasing of the strength of the interaction of a particle, therefore the probability of interaction, and as a consequence also its volume of occupation in the phase space, that is, the mass. Indeed, doubling the charge means logarithmically doubling, i.e. squaring, the interaction probability, $P \propto \alpha \propto g^2$. Since in the present case we increase $|Q|$ by $\frac{1}{3}$ of the unit electric charge, we expect that, in passing from the electron to the lightest quark, besides the factor 6.66, we approximately gain an extra $(\alpha_\gamma)^{-1/3}$ factor [56].

[55] Notice that the usual charge assignment breaks the $SU(2)$ symmetry explicitly.

[56] No further 1/3 normalization factors are needed, because in this operation we are leaving unchanged the $SU(3)$ indices.

The upper quark of the $SU(2)$ pair passes on the other hand from $|Q| = \frac{2}{3}$ to $|Q| = \frac{1}{3}$, but it does not acquire mass shifts (in the sense either of expansion or of contraction of its volume in the phase space) other than what already inherited by the expansion in the phase space of the lower partner quark. The two are in fact separated by an $SU(2)$ rotation, and the absolute value of their mass difference remains the same: the electric charge modification $|Q| \, : \, \frac{2}{3} \to \frac{1}{3}$ has to be seen as the result of an $SU(2)_{\text{w.i.}}$ rotation from the lower member of the pair, therefore a $|\Delta Q| = 1$ rotation, not as a charge shift by $|Q| = \frac{1}{3}$. If, in order to better compare with experimental data, instead of using the inverse of 6.85, we consider the current value of the fine structure constant at the MeV scale we will obtain in section 7.4, putting everything together we get:

$$\begin{aligned} m_d &\approx 4,39\,\text{MeV}\,, \qquad (7.10)\\ m_u &\approx 2,50\,\text{MeV}\,, \qquad (7.11)\end{aligned}$$

so that:

$$\delta m_{u/d} = m_u - m_d \approx 1,89\,\text{MeV}\,. \qquad (7.12)$$

7.1.3 *The charged particles of the second family*

The masses of the charged particles of the second family are obtained from 6.68, 6.69, 6.70. At present time, they are:

$$\begin{aligned} m_\mu &\approx 94\,\text{MeV}\,; \qquad (7.13)\\ m_s &\approx 167\,\text{MeV}\,; \qquad (7.14)\\ m_c &\approx 1,539\,\text{GeV}\,. \qquad (7.15)\end{aligned}$$

7.1.4 *The charged particles of the third family*

The masses of the charged particles of the third family are obtained from 6.71, 6.72, 6.73:

$$\begin{aligned} m_\tau &\approx 13,85\,\text{GeV}\,; \qquad (7.16)\\ m_b &\approx 56\,\text{GeV}\,; \qquad (7.17)\\ m_t &\approx 2749\,\text{GeV}\,. \qquad (7.18)\end{aligned}$$

One can see that, up to the second family, the mass values, although all more or less slightly differing from those experimentally measured, are anyway of the correct order of magnitude. The values obtained for the third family, instead, seem to be hopelessly wrong. In the next sections we will discuss how these "bare" values get corrected by a refinement in our approximation.

7.2 *Corrections to masses*

Some of the mass values we have obtained are close to the experimental ones. Other masses, in particular those of the charged particles of the third family, are decidedly out of their

experimental value by almost one order of magnitude. In any case, no one really coincides with its known experimental value. Indeed, as we already pointed out, free elementary particles correspond to a conceptual classification of the real world, that makes sense only in the case of weakly coupled states. In our scenario, for the leptons this condition is better and better satisfied as the Universe expands. Quarks are instead strongly coupled, and for us their coupling will become stronger and stronger as time goes by. In order to disentangle the properties of the elementary states as "free states", we have mapped to a logarithmic representation of the string vacuum. In this picture, owing to the linearization of the string space, it was easier to consider the ratios of the volumes occupied in the phase space by the various particles, and their interactions. However, as we pointed out, this mapping works only at times close to the Planck scale, where the "logarithmic world" becomes weakly coupled. At a generic finite time, the entire spectrum of particles is "strongly coupled": not only the "colour" interactions are strong, but even the electro-weak symmetry can hardly be expanded around a vanishing value of the coupling. As we discussed in section 6.1.2, in such a world, the only true "asymptotic" state is neutral to all the interactions. We have identified this as a bound state made of neutron, proton, electron and its neutrino and their antiparticles. The corresponding mass scales as $\mathcal{T}^{-3/10}/2$. Strictly speaking, at finite time this is the only true "bare" state of our theory, and its mass scale can be used in order to set the scale of the universe. The problem of correctly computing masses is then twofold:

1) first of all there is the fact that masses, as they are experimentally derived, correspond to a theoretical framework of scale-running, finite volume regularization of a theory basically defined in an infinitely extended space-time. The values we gave in the previous sections must be corrected in such a scheme, taking as starting point the "regularized" value of the fundamental scale, related to the neutron's mass;

2) besides this correction, we must also consider that the masses below the $\mathcal{T}^{-3/10}$ scale should be treated as perturbations of this scale. This is the case of the proton and the neutron, which are made of up and down quarks of the first family, but have a mass much closer to the GeV scale than to the one of the quarks they are made of. By consistency, we should apply the same argument also to the electron. The electron's mass too should be considered as a perturbation of the mean scale. And indeed, strictly speaking it is: free electrons exist for a short time, until they "recombine" into atoms or anyway they bound into some materials. However, since the strength of their interaction (as well as that of neutrinos) decreases with time and at present is sufficiently small, we can safely speak of electrons as free states;

3) for masses above the $\mathcal{T}^{-1/3}$ scale, things are reversed: it is $\mathcal{T}^{-1/3}$ which is rather a perturbation of the bare mass scale.

When we consider the corrections to the masses we want to compute, it is therefore of primary importance to distinguish, at least from an ideal point of view, what are the corrections with respect to what: with respect to the mass of a stable state, or to that of an unstable phase, whose existence can anyway be indirectly detected? This problem is of particular relevance when we talk about quarks. For instance, the "bottom" or "top" mass. In this case, what are indeed measured are the masses of the quark compounds, mesons that exist for a short time: transitory states, that we mostly know through their decay products.

It is currently assumed that the mass of the compound reflects with a good approximation the mass of the heavy quark. However, for us the question is: how is this mass related to the "bare" mass quoted in 7.17, 7.18?

7.3 Converting to an infinite-volume framework

As discussed in section 6.2.9 for the couplings, also for elementary masses, in order to convert their values to on-shell values at the appropriate scale, in a framework of infinitely extended space-time, we must treat them in a renormalization scheme based on a finite-volume regularization. In this picture they are converted to running values, fixed to reduce to the "bare" values at a certain age $\mathcal{T}$ at the fundamental mass scale for spinors, the $\mathcal{T}^{-1/2}$ scale. The "regularization prescription" is that the effective value of the age of the universe $\mathcal{T}$ is adjusted on the neutron mass through its relation to the $\mathcal{T}^{-3/10}$ scale 6.10.

The evaluation of masses proceeds therefore along a sequence of perturbative steps: at first we roughly determine, as in sections 7.1.1–7.1.4, the energy scale "at rest" of the a certain particle. Then we improve the computation by letting the mass to run from the fundamental $\mathcal{T}^{-1/2}$ scale to the specific scale, obtaining thereby an improvement in the perturbation process. Exactly knowing the running of masses entails a detailed knowledge of the interaction and decay processes the particle is involved in: these in fact decide what is the weight of a particle in the phase space. This investigation too can be viewed as part of a sequence of perturbative steps. At the first step, the logarithmic running of masses can be inferred from 6.46, whose differentiation produces a renormalization group equation for the running of mass ratios as opposite to the running of couplings. In this case, the coupling concerned is $\alpha_{SU(2)_{\Delta m}}$, that, according to 6.84, runs more or less like the electromagnetic coupling α_γ, just a bit slower. It is a kind of "non-chiral weak coupling", and the correct evaluation of its beta function suffers of the same theoretical problems of the other gauge couplings, namely of the lack of exact knowledge of the most appropriate effective theory (non-supersymmetric? minimally supersymmetric? with an effective Higgs field?...). In section 6.2.9 we assumed that, in first approximation, in the effective representation of the physical configuration, couplings run logarithmically with an effective beta-function such that, starting from their "bare" value at the actual $\mathcal{T}^{-1/2}$ scale, they meet at zero at the Planck scale. We can here assume that this holds for the $\alpha_{SU(2)_{\Delta m}}$ coupling too. From 6.46 we derive then that the relative variation of a mass along a certain scale variation is opposite to the one of the $SU(2)_{\Delta m}$ coupling:

$$\frac{\Delta m}{m} = -\frac{\Delta \alpha_{SU(2)_{\Delta m}}}{\alpha_{SU(2)_{\Delta m}}} \,. \tag{7.19}$$

Notice that, while the inverse couplings decrease to zero, therefore couplings increase when going toward the Planck scale, masses decrease. This is correct, because what we are giving here are relative corrections to mass ratios, not masses in themselves. It must be kept in mind that this linearized representation makes only sense reasonably away from the Planck scale. In first approximation the mass corrections are of order:

$$\frac{\Delta m_i}{m_{\rm i}} \approx \frac{\ln \mu_0 - \ln \mu_i}{\ln \mu_0} \,, \tag{7.20}$$

where μ_i are the mass scales given in sections 7.1.1–7.1.4, $\mu_0 = \left(\frac{1}{2}\right)^2 \mathcal{T}^{-1/2}$, μ and μ_i are expressed in reduced Planck units, an appropriate Planck mass rescaling in the argument of each logarithm being implicitly understood. Indeed, since masses are obtained from the expressions of mass ratios, the higher mass of a pair is obtained as a function of an inverse coupling times the lower mass, which sets the scale of the process. Effectively, expression 7.20 is therefore shifted to:

$$\frac{\Delta m_{\mathrm{i}}}{m_{\mathrm{i}}} \approx \frac{\ln \mu_0 - \ln \mu_{\mathrm{i}} + \ln \mu_{\nu_e}}{\ln \mu_0} \,. \tag{7.21}$$

The first neutrino mass remains unvaried. For the other masses, we obtain:

$$m_{\nu_\mu} : \quad 2,945\,\mathrm{eV} \;\rightarrow\; 2,739\,\mathrm{eV}\,; \tag{7.22}$$

$$m_{\nu_\tau} : \quad 4,3385\,\mathrm{KeV} \;\rightarrow\; 3,731\,\mathrm{KeV}\,; \tag{7.23}$$

$$m_e : \quad 0,639\,\mathrm{MeV} \;\rightarrow\; 0,505\,\mathrm{MeV}\,; \tag{7.24}$$

$$m_\mu : \quad 94\,\mathrm{MeV} \;\rightarrow\; 67,7\,\mathrm{MeV}\,; \tag{7.25}$$

$$m_\tau : \quad 13,85\,\mathrm{GeV} \;\rightarrow\; 8,99\,\mathrm{GeV}\,; \tag{7.26}$$

$$m_u : \quad 2,50\,\mathrm{MeV} \;\rightarrow\; 1,93\,\mathrm{MeV}\,; \tag{7.27}$$

$$m_d : \quad 4,39\,\mathrm{MeV} \;\rightarrow\; 3,35\,\mathrm{MeV}\,; \tag{7.28}$$

$$\delta m_{u/d} : \quad 1,89\,\mathrm{MeV} \;\rightarrow\; 1,42\,\mathrm{MeV}\,; \tag{7.29}$$

$$m_c : \quad 1,539\,\mathrm{GeV} \;\rightarrow\; 1,048\,\mathrm{GeV}\,; \tag{7.30}$$

$$m_s : \quad 167\,\mathrm{MeV} \;\rightarrow\; 118,9\,\mathrm{MeV}\,; \tag{7.31}$$

$$m_t : \quad 2749\,\mathrm{GeV} \;\rightarrow\; 1582\,\mathrm{GeV}\,; \tag{7.32}$$

$$m_b : \quad 56\,\mathrm{GeV} \;\rightarrow\; 35,3\,\mathrm{GeV}\,. \tag{7.33}$$

The only elementary particle mass we can here use for a precise comparison with experimental data is the one of the electron: neutrino masses are not yet known, and the other masses will undergo further corrections (see next sections).

The correction 7.24 must be considered as a first order correction: once determined at "order zero" the bare mass, $0,639\,\mathrm{MeV}$, we have rescaled it according to 7.21, by recalculating the effective coupling on the zero order electron's scale. Now that we have the first order electron's mass scale, $\sim 0,505\,\mathrm{MeV}$, we can improve our approximation by recalculating the effective coupling on this new scale, and using this newly obtained relative mass correction in order to correct the scale of the $0,639\,\mathrm{MeV}$. We obtain in this case:

$$m_e|_{2^{nd}}\,;\; 0,505 \;\rightarrow\; 0,5069397\ldots\,\mathrm{MeV}\,. \tag{7.34}$$

This is still about 1% lower than the experimental value. Indeed, in order to get the *physical* mass of the electron, to the "bare" mass 7.24 we must add also the masses of the lighter states. The reason is the following. In the derivation of the mass ratios of section 6.2.3, namely proceeding from 6.46, there is the implicit assumption that all lighter masses of a particle belong to a subspace of its phase space. Suppose we have just two particles, particle A with mass m_A, and particle B, with mass $m_B = \alpha\, m_A$, $\alpha < 1$. When we say that α is the ratio of the two volumes in the phase space, we also imply that particle A is heavier than particle B in that the space of B has been obtained by a process of symmetry reduction,

by truncating the space of A. Particle A has more interaction/decay channels than B, because the space of A contains the space of B. Let's now consider the full phase space of a sub-universe consisting of A and B. The full volume is:

$$V(A) + V(B) = V(A) + \alpha V(A) . \tag{7.35}$$

Now, in our specific case A is the electron, and B is basically the τ-neutrino (we neglect here the other neutrinos, that give corrections of order $\mathcal{O}(\alpha^2)$). When we measure the mass of the physical electron, what we look at is the modification to the geometry of the space-time produced by the existence of the electron. For what we just said, deriving the electron's mass from 6.46 implies considering that, when generating the electron, we generate also the τ-neutrino and the ligher particles. They also interact, and the modification to the whole phase space produced by the existence of the electron is indeed the full $V(A) + V(B) = V(A) + \alpha V(A)$. This implies that what we call the physical electron mass is the sum of the bare electron mass 7.24 *plus*, in first approximation, the mass of the τ-neutrino. Summing to 7.34 the ν_τ mass 7.23, we obtain then:

$$m'_e|_{2^{nd}} ; \; 0,50694 \;\to\; 0,51057 \ldots \text{MeV} . \tag{7.36}$$

Of course, we can correct to the second order also the ν_τ mass, and further refine our evaluation. At this order the ν_τ mass gets increased, thereby increasing also the estimate of the electron's mass. A further recalculation of the coupling at the new scales leads on the other hand to a subsequent lowering of all masses. The approximation of the electron's mass proceeds through a converging series of "zigzag" steps of decreasing size, below and above the final value. One can easily see that in this way we better and better approximate the experimental value of the electron's mass (see [58]). However, we don't want here to go into a detailed fine evaluation of mass values, because $\sim 1\% \div 0,1\%$ is our best precision in many steps of our analysis of masses.

In general, accounting for the shifting of phase space 7.35 amounts in a small ($\mathcal{O}(\alpha^{-1})$) correction to mass values, but for the quarks of the first and second family the relative change is much higher ($\mathcal{O}(\sqrt[3]{\alpha^{-1}})$ and $\mathcal{O}(\sqrt{\alpha^{-1}})$ respectively). Once this is taken into account, the masses of the up and down quarks get further corrected to:

$$m_u : \quad 1,93\,\text{MeV} \to 2,435\,\text{MeV} ; \tag{7.37}$$

$$m_d : \quad 3,35\,\text{MeV} \to 5,785\,\text{MeV} ; \tag{7.38}$$

$$\delta m_{u/d} : \quad 1,42\,\text{MeV} \to 3,35\,\text{MeV} . \tag{7.39}$$

7.4 The fine structure constant: part 2

Let's now come back to a more precise determination of the fine structure constant. As discussed in section 6.2.9, the fine structure constant is the value of α_γ^{-1} at the electron's scale, the scale that can be considered as the reference for the operational definition of the electric charge. According to our analysis of section 6.2.3, summarized in the diagram 10, the phase space of the elementary particles divides into two subspaces, the electrically uncharged and the charged space, the latter being the upper one in the sense that all charged particles

are heavier than the uncharged ones. This second subspace starts at the electron's scale. As we discussed at page 131, section 7.1.2, after the up-down flip in the quarks of the first family, the phase space gets further expanded by a $\sqrt[3]{\alpha_\gamma^{-1}}$ factor. This shift modifies the effective strength of the projections applied in order to get the mass hierarchy of section 6.2.3 in the sub-volume of the phase space corresponding to the first charged family. As a consequence, it modifies also the effective weight of the corresponding states, and the ratio of the effective $U(1)_\gamma$ and the $SU(2)_{\Delta m}$ beta-functions *around this scale*. The effect is that, as the states weight more, the effective running of the coupling is faster, or, equivalently, the one of its inverse slower. Namely, as the volumes of the matter phase space are expanded (or, logarithmically, shifted), the value of the electromagnetic coupling at the scale m_e effectively corresponds to the value of the coupling *without correction* at a run-back scale, $m_e^{\text{eff.}}$. The amount of running-back in the scale of the logarithmic effective coupling is equivalent to the amount of the forward shift in the logarithmic representation of the volumes of particles in the phase space. If volumes get multiplied by a factor, their logarithm gets shifted, and so gets shifted back the scale at which the coupling in its logarithmic representation is effectively evaluated. From an effective point of view, we can therefore derive the value of the fine structure constant by evaluating the electromagnetic coupling proceeding as in 6.94, but at a scale a factor $\sqrt[3]{\alpha_\gamma^{-1}}$ below the electron's scale, rather than precisely at the electron's scale as we did in 6.97 (see also Appendix D). To have a first rough estimate, we can use 6.97) to calculate that the effective scale μ_γ is lower than $0,511\,\text{MeV}$ by a factor $\sim 5,102549027\ldots$. In this case we obtain:

$$\alpha_\gamma^{(1)\,-1}|_{m_e} \;=\; 137,0700548\,. \tag{7.40}$$

In order to improve our evaluation, we need a better approximation of the shape and size of the effective shift of the phase space of the first family. If we consider that the $\sqrt[3]{\alpha_\gamma^{-1}}$ shift on the up quark translates also to the down quark, the heavier in this case, we should conclude that the scale at which to evaluate $\sqrt[3]{\alpha_\gamma^{-1}}$ is around the down quark mass scale. Using the value 7.28 for the point of evaluation, we obtain:

$$\alpha_\gamma^{(2)\,-1}|_{m_e} \;=\; 137,0366167\,. \tag{7.41}$$

In order to further improve the estimation, one should then proceed as we did for the electron, by iterated steps of corrections of the down and electron scale, recalculating the $\alpha_{SU(2)}$ factors at the new scales to obtain improved estimations of m_d and of $\alpha_\gamma^{(0)}$ at the down mass scale, and so on, obtaining a series of converging "zigzag" steps. The first step corresponds to a slight increasing of the effective down mass, thereby lowering the factor $\sqrt[3]{\alpha_\gamma^{(0)\,-1}}$, eventually resulting in a slight, higher order decrease of the value of the inverse of the fine structure constant. The value 7.41 is already less than $0,001\,\%$ above the experimental one [58], and these considerations induce to expect that the further steps of the approximation do improve the convergence toward the experimental value. However, one should not forget the major point of uncertainty, namely that we are here attempting to parametrize the effective modification of the size of the projections applied to the phase space, and therefore the value of the fine structure constant, due to a local dilatation of the phase space. A true fine evaluation of α_γ^{-1} requires first of all a better approximation of this effect. Last but not least, there is the question whether, and at which extent, a convergence toward the

official "experimental value" of this parameter should be expected and desired. Beyond a certain order of approximation, current evaluations heavily rely on QED techniques, and are extrapolated within a theoretical scheme that only at the first orders corresponds to the one discussed in this paper . A mismatch beyond this regime of approximate correspondence does not necessarily implies and indicates that the values here obtained are wrong, provided the effective computation of physical amplitudes nevertheless produces correct results.

Finally, we repeat and stress that, in our framework, the electric charge is time-dependent, and 7.41, possibly corrected at any desired order, only represents the present-day value of this parameter. The rate of the time variation at present time can be easily derived from the very definition. From 6.80 and 6.84 we obtain:

$$\frac{1}{\alpha}\frac{d\,\alpha}{d\,t} \;=\; \frac{1}{28} \times \frac{47}{45} \times \frac{1}{\mathcal{T}} \,. \tag{7.42}$$

In one year, the expected relative variation is therefore of order $\approx\, 3 \times 10^{-12}$. This is a rather small variation, however not so small when compared with the supposed precision with which α is obtained. Indeed, the most recent measurements give for its inverse a number with precisely 12 digits, a number whose variation could be observed by repeating the measurement at a distance of some years. Since however a fine experimental determination of α depends, through the theoretical framework within which it is derived, on time-varying parameters such as lepton masses etc..., it would not be an easy task to disentangle all these effects to get the "pure α time-variation". This kind of effects can be better detected when expanded on a cosmological scale, as we will discuss in section 10.1.

7.5 The Heavy Mass Corrections

Any perturbative approach is based on the identification of a "bare" configuration, a set of states, which serve as starting point for a series of corrections. For the sake of consistency, these must be "small" as compared to the main contribution, the "bare" quantity. The selection of a "bare" set is in general not uniquely determined: precisely in string theory, a complete investigation requires a full bunch of "dual constructions", built as perturbations around different sets of bare states. In our case specific case, several physical quantities can be viewed both from the the physical ("exponential") picture, and from the point of view of a logarithmic representation of the vacuum they are mapped to. As we have discussed, a comparison with experimental data presented within a framework of infinite space volume and free particles may require to map quantities to this picture. The advantage of a logarithmic representation resides in that, in the cases of interest, the logarithm maps large quantities into small ones, and vice-versa. In particular, in the case of masses which are below the Planck scale, since everything is measured in units of the Planck scale, higher masses are mapped into smaller ones:

$$m_1 \,>\, m_2 \qquad \Rightarrow \qquad |\ln m_1| \,<\, |\ln m_2| \,, \qquad (|m_1|\,,\,|m_2| \,<\, 1\,) \,. \tag{7.43}$$

When we evaluate the corrections to the masses, we must therefore consider the perturbations to the bare values as seen from both the point of view of the "exponential picture", where

for instance the mass of the up quark is smaller than the mean mass scale $m_{3/10}$, which has then to be considered as the "bare" scale around which to perturb, and the point of view of the "logarithmic picture", in which, according to 7.43, this hierarchy is reversed, and it is the quark mass which is going to be perturbed. Once resummed, depending on the picture one considers, the mass corrections may therefore be quite large, larger than the scale they are going to correct.

We will in the following consider two types of corrections to the bare quark masses, depending on whether they refer to a perturbation of the stable non-perturbative mass scale, the $m_{3/10}$ (the case of the proton mass), or to unstable particles, whose existence in itself is a perturbation of the mean configuration of space-time.

7.5.1 Stable particles

At large volume-age of the universe, $1 \ll \mathcal{T} \to \infty$, electro-weak interactions are very weak, while strong interactions become stronger and stronger. Asymptotically, the universe settles to a configuration in which only the lightest particles of the decay chain are normally present, the probability of producing higher mass, unstable ones becoming lower and lower. The particles charged under the strong interaction tend to form bound states, "attracted" by the mean mass scale "$m_{3/10}$", defined in eq. 6.10. In practice, this means that the universe tends toward a world with matter made out of up and down quarks, electrons and neutrinos [57]. The masses of these objects as *free* particles are those computed in sections 7.1.1, 7.1.2, and corrected in 7.22–7.33. Quarks however are not present as free particles, but form the bounds we call proton and neutron, which are stable in the sense that the neutron's decay into proton+electron+neutrino is balanced by the inverse process of electron and neutrino capture by the proton. As discussed, the stable mass scale $m_{3/10}$ corresponds to the rest energy of this system, namely the neutron + proton-electron-neutrino plus their antiparticles. If we look at the masses of the quarks and leptons constituting this system, we see that the neutron is made of an heavier quark set than the proton, and that the quark mass difference between neutron and proton is higher than the sum of the electron and neutrino masses. This means that the neutron decay into proton+electron+neutrino leaves some amount of energy. As far as the $m_{3/10}$ compound at equilibrium is concerned, this energy must be included in the account. This allows us to deduce that "at equilibrium" $(1/2)\, m_{3/10}$ corresponds to four times the mass of the neutron. Furthermore, neutron and proton differ for their electromagnetic charge, i.e. for "weak" interaction properties as compared to the strong coupling; it is therefore reasonable to expect that their mass difference is basically due to the mass difference between the up and down quark. On the other hand, the $m_{3/10}$ scale is much higher than the down-up mass difference, and, as it corresponds to a stable scale, it can be considered weakly coupled. This impplies that the quark mass difference can be treated as a small perturbation of the $m_{3/10}$ scale. We can therefore write:

$$m_{\rm p} \approx \frac{1}{4}\, m_{3/10} \, + \, \mathcal{O}(\delta m_{u,d}) \, . \tag{7.44}$$

[57] We will comment later about muon- and tau- neutrinos.

Indeed, since the proton is a stable particle, the difference in the volume occupied in the moduli space by neutron and proton is entirely due to the difference of the volumes occupied by the quarks they are formed of. Differently from what discussed at pages 7.3 and 7.3, a correction to the down quark mass does not require summing the mass of the lighter states, as it was the case for instance of the electron, whose bare mass had to be corrected by adding the ν_τ mass. The reason is that, in this case, once bound to form the heavy, strong-coupling-singlet compound, the lighter particle, the quark up, does not interact anymore: it does not have an independent phase space, as it was the case of the τ-neutrino. For what concerns the physical mass of the proton and the neutron, namely, as long as, like for the case of the electron, we look at the modification caused to the geometry of space-time by the existence of the proton and the neutron, the phase space of the up quark does not add to the phase space of the "bare" down quark. The quark mass difference entering in this game is therefore the (corrected) bare quark mass difference 7.29. *At the quark scale*, this is $\delta m_{u/d} = 1,42\,\mathrm{MeV}$. However, for what matters the mass difference between neutron and proton, the scale at which the quark masses have to be run is the proton/neutron scale. At this scale, once recalculated according to 7.21, the quark masses are:

$$m_u|_{E=m_n} = 1,7189\,\mathrm{MeV} \qquad (7.45)$$

$$m_d|_{E=m_n} = 3,0183\,\mathrm{MeV} \qquad (7.46)$$

that imply:

$$m_d - m_u|_{E=m_n} = 1,299\,\mathrm{MeV} \approx m_{\mathrm{n}} - m_{\mathrm{p}}\,, \qquad (7.47)$$

quite in good agreement, apart the usual $\mathcal{O}(1\,\%)$ mismatch, with the experimental value of the neutron-proton mass difference [58]. Notice that, in their logarithmic running, masses, and mass differences, decrease when increasing the scale. This is the opposite of what happens in the real, cosmological scaling. The point is that in the real scaling they all tend to 1 in Planck units. As it happens for the couplings, also masses tend to the logarithm of 1, namely, to zero, at the Planck scale.

Expression 7.44 accounts with good approximation for the behaviour of the proton-to-neutron mass relation far away (i.e. well below) the Planck scale. As we get close to this scale, this approximation looses its validity.

7.5.2 *Unstable particles*

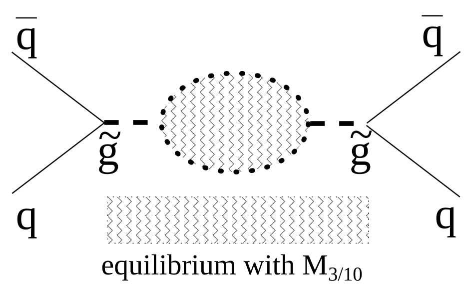
Let's now consider the particles that exist only for a short time: the leptons μ, τ and the mesons. As seen from the point of view of the universe along the running of its history, at a late stage, as it is today, the existence, i.e. the production and decay, of these particles can be seen as a fluctuation out of a "vacuum" characterized by the mass scale $m_{3/10}$, of which they are a perturbation. Indeed, since we are going to compare masses with values given in an infinite-volume theoretical frame corresponding to a logarithmic picture, in practice it is the lower mass what is going to be seen as the "bare" value to be corrected. This may be the $m_{3/10}$ mass itself in the case of particles with a bare mass higher than the $m_{3/10}$ scale, as is the case of the particles of the third family (the quarks top, bottom and the τ). Or it can

be the mass of the particle, as is the case of the second family (charm, strange and muon). In any case, the correction is of the form:

$$M_0^2 \; \rightsquigarrow \; M^2 \; \approx \; M_0^2 \left(1 + \alpha \, \frac{m^2}{M_0^2} \right) \,, \tag{7.48}$$

where M_0 and M are the bare and the corrected mass, and m is the perturbing mass, the mass scale with which the state of mass M is in contact through an interaction with strength $\alpha = g^2 / 4 \pi$.

In the case of particles of the third family, M is the $m_{3/10}$ scale, that from the point of view of a logarithmic picture is the higher scale, and m the bare quark or lepton mass. For the second family, things work the other way around: M is the bare mass of the particle, and m corresponds to the $m_{3/10}$. When we say "the bare quark mass" here we intend something different from the usual concept of bare quark mass. As they are usually given, quark masses are in general directly derived from the mass of mesons they form, possibly subtracted of the mass of the partner quark they are bound with, and corrected within the framework of an $SU(3)$-colour symmetry based model of hadrons. Apart from the case of the up and down quarks, the quark mass turns out to be, although not really coinciding and sometimes considerably different [58], anyway of the same order of magnitude of the meson mass. In our case, bare mass means instead the value given in 7.27–7.33. The coupling α is in general the electromagnetic coupling, which provides the strongest interaction between the two mass scales. Masses enter in expression 7.48 to the second power, because this mass correction can be viewed as a propagator correction of an effective boson, as here illustrated



where q and $\bar{q}$ stand for a quark-antiquark pair, in the simplest case for instance in a π-meson. Indeed, an expression similar to 7.48 could be considered also for the stable barions considered in section 7.5.1. In that case, the strongest contact between the two scales, the up and down quark scale and the $m_{3/10}$ scale, is given by the strong coupling itself, of order one. The correction ends up therefore to the $m_{3/10}$ scale itself. For the π-mesons, or the other mesons, K, C, B etc... (these last ones more or less "by definition" in direct relation to the mass of their heaviest quark), although their constituents interact strongly, this interaction involves the quarks within each meson. The strongest contact between the two scales is

[58] See for instance the case of the strange quark and the K mesons.

however given by the electromagnetic interaction, and α is basically the electromagnetic coupling.

In the case of neutrinos, their only contact with the $m_{3/10}$ scale occurs through the weak coupling. In itself, $\alpha_{\text{w.i.}}$ is even a bit stronger than the electromagnetic coupling. However, the effective strength of the interaction is of order:

$$\alpha^{\text{eff.}} \approx \alpha_{\text{w.i.}} \times \frac{m_\nu^2}{M_W^2}, \tag{7.49}$$

where $\alpha^{\text{eff.}}$ already takes into account typical energies of neutrino processes, and should not be confused with G_F, the Fermi coupling constant. The neutrino mass corrections are therefore extremely suppressed.

The correction 7.48 reduces the effective mass of the t or b quarks and the τ lepton by around one order of magnitude, producing values close to those experimentally measured. More precisely, the top mass gets corrected to:

$$m_t \to \sim 164\,\text{GeV}, \tag{7.50}$$

where, besides the value 7.32, we have also used the value of the electromagnetic coupling logarithmically corrected to the bare top scale (α_γ^{-1} : $183,78 \to 92,91$) [59]. As in the case of the electron, this value too should be corrected at higher orders, by recalculating the "bare" top mass, from the 1582 GeV of the first order, to a second order value, to be used as starting point for the correction, to be plugged in 7.48. Then, as we did for the electron, in order to catch the full phase space of the physical top particle, we must add the lighter masses, the heaviest of which are the bottom, tau, and charm masses. Then, here too one can easily see that these higher order corrections better and better approximate the experimental value of the top mass. Let's see the first steps of this correction. First of all, we recalculate the relative mass correction, or equivalently the relative coupling correction, run at the new corrected top mass, 1582 GeV. We obtain:

$$m_t^{(0)} = 2749\,\text{GeV} \to 2749 - (2749 \times 0,4167) = 1603,5298\,\text{GeV}. \tag{7.51}$$

To this, we must sum the non-negligible contributions of the bottom, τ and charm bare masses, obtaining:

$$m_t' \approx 1603,53 + 35,3 + 8,99 + 1,048 = 1648,87\,\text{GeV}. \tag{7.52}$$

Of course, to be more precise we should re-correct at the second order also the bottom, τ and charm masses, something we are not doing here. Re-plugging 7.52 in 7.48, we obtain:

$$m_t'' \approx 171,07\,\text{GeV}, \tag{7.53}$$

[59] In principle, this value could be affected by the shift in the effective beta function, centered to the electron's scale, we discussed in section 7.4. However, we don't have a recipe in order to derive the full non-linear effective running of the electromagnetic coupling. We suppose that the local modification has its peak at around the electron/up/down scale, and tends to vanish both toward the $\mathcal{T}^{-1/2}$ and the m_t scale. Therefore, we neglect it in this and the following computations, already affected in themselves by possibly larger uncertainties.

quite more in agreement with the experimental value, which is around $\sim 171,4 \pm 1,7\,\text{GeV}$ [68]. We don't go further in the refinement of 7.53, because, to start with, we should recalculate also the bottom, τ and charm bare masses. Then, to be more precise, we should also take into account the modifications to the effective coupling and bare mass logarithmic scales, as due to the $SU(3)$ normalization factors of the quark mass ratios, 1/3 and 1/9 for the bottom and the top of each $SU(2)$ doublet. All these corrections contribute for at most an order $\sim 1\,\%$, therefore an uncertainty lower than the error in the experimental value of the top mass. More importantly, we must warn here that the agreement we obtain between our estimate and the experimental value has to be taken more as the indication of the plausibility of our analysis, rather than a real fine test. We are trying to evaluate the ratios of the volumes in the phase space of the particles in a rather complicated part of the spectrum, where the regions of validity of our simple perturbative dual approaches meet. For instance, it is not completely clear whether the best approximation is obtained by summing to the top phase space the lower masses *before* the correction through the $m_{3/10}$ scale, or *after* it. Here and in the following we choose the first option. In the case of the top quark, since the top scale is well above all these scales, this does not make such a big difference. Things become however more critical when looking at the corrections to the lower masses, such as the one of the bottom quark, the τ or the charm quark.

For the bottom, the effective coupling we use is the inverse electromagnetic at the bottom scale, $\alpha_\gamma^{-1}|_b \sim 102,95$. We obtain:

$$m_b \; \to \sim 3,61\,\text{GeV}\,. \tag{7.54}$$

This scale too should then be corrected in a way similar to the top mass. Adding the tau and charm masses, we obtain:

$$m_b \; \to \sim 4,57\,\text{GeV}\,. \tag{7.55}$$

This value is slightly above the average experimental estimate. However, the latter is basically *extrapolated* from the B-meson width, and 7.55, although above the extrapolated value, is actually still compatible with the mass of the B-meson. A serious comparison would require a better understanding of the theoretical uncertainties underlying the entire derivation, both on the side of our evaluation of volumes in the phase space, and on the side of the experimental derivation: for consistency, the extrapolation from experimental data should be done entirely within the light of our theoretical scheme.

For the τ lepton we use a value of the electromagnetic coupling run to the lepton's scale, $\alpha_\gamma^{-1}|_\tau \sim 106,55$, and obtain:

$$m_\tau \; \to \sim 1,28\,\text{GeV}\,. \tag{7.56}$$

For the further corrections to this value, analogous arguments apply also here, with the difference that, being the τ mass so close to the $m_{3/10}$ scale, the final result is more sensitive to these corrections than in the top and bottom case. For instance, at the second order the corrected bare τ mass, instead of 7.26, is $m_\tau|_{2^{nd}} \sim 9,52\,\text{GeV}$, that gives $1,32\,\text{GeV}$. Adding the charm mass, we get a further correction by some 5%, leading to:

$$m'_\tau \; \sim \; 1,39\,\text{GeV}\,. \tag{7.57}$$

As it is also the case of the quarks of this family, in particular the bottom quark, it doesn't make however sense to go on with refinements of scale evaluations, as it is already clear that something more fundamental is here missing, in order to explain the gap between the values we obtain and the so-called experimental one ($\sim 1,78\,\mathrm{GeV}$ [58]). As we said, a better understanding of the corrections to the volumes of phase spaces around the $m_{3/10}$ scale for unstable particles is in order. In the case of the bottom quark, the experimental value too is strongly affected by model-dependent considerations, and things are even more complicated. I hope to report some progress in the future.

When we pass to the second family, analogous considerations hold for the charm quark, whose mass is extremely close to $m_{3/10}$. In first approximation, by inserting the renormalized value of the electromagnetic coupling at the bare charm mass scale, $\alpha_\gamma^{-1}|_c \sim 113,5$, we obtain a slight decrease of the quark mass:

$$m_c \,:\, 1,048 \to 0,946\,\mathrm{GeV}\,. \tag{7.58}$$

However, as it is already evident from the τ mass evaluation of above, as the bare scale approaches the $m_{3/10}$ scale, our perturbation method starts showing its limitations. Indeed, in the case of the charm quark, it would be also possible to invert the role of bare mass and perturbing mass, using the charm bare mass 7.30 as the mass M in the expression 7.48, and, for m, the neutron mass, obtaining:

$$m^{\prime}_c \,:\, 1,048 \to 1,051\,\mathrm{GeV}\,. \tag{7.59}$$

Including the strange-quark mass shift, we would obtain a light increase to:

$$m^{\prime}_c \,\sim\, 1,170\,\mathrm{GeV}\,. \tag{7.60}$$

Similar considerations as for the bottom and τ masses are in order here too, and we leave any further analysis for the future.

For the strange quark and the μ-lepton, they are below the $m_{3/10}$ scale, and, as we start to get far away from it, the reliability of our estimate starts to improve again. For the strange quark, we use $\alpha_\gamma^{-1}|_s \sim 117,94$, to obtain, if we don't consider the μ-mass shift:

$$m_s \,\to\sim 147\,\mathrm{MeV}\,, \tag{7.61}$$

and, when including the muon mass shift:

$$m_s \,\to\sim 205,7\,\mathrm{MeV}\,. \tag{7.62}$$

A comparison with what is known as the experimental value of the strange quark mass is affected by theoretical considerations. In itself, the strange quark mass is extrapolated via $SU(3)_c$-related techniques from the width of the K-mesons. Surely, in the space of the K-mesons there is also the μ- channel. However, when the "bare" s-quark mass is disentangled from the total width, does this mean that also the μ- shift gets decoupled? In this case, the value to be considered for a comparison should not be the second one, 7.62, but the

μ-unshifted one, 7.61. The difficulties rely here also on the fact that we are comparing *extrapolated* values, not true "experimental" ones.

Finally, for the muon we use $\alpha_{\gamma}^{-1}|_{\mu} \sim 119,42$, that leads to:

$$m_{\mu} \; \to \sim 109,4\,\mathrm{MeV}\,, \tag{7.63}$$

and, when including the electron mass shift:

$$m_{\mu} \; \to \sim 109,8\,\mathrm{MeV}\,. \tag{7.64}$$

One may notice that our mass corrections become the less and less precise as we get closer to the $m_{3/10}$ mass scale. Indeed, our approximation of the correction works better when the bare scale of the particle is far away from $m_{3/10}$, so that we can either treat the particle's scale, or the $m_{3/10}$ scale, as the perturbing or the perturbed scale. When they are close, other "non-linear" effects become important, and with our approximation we systematically obtain an overestimate for the particles with a mass below $m_{3/10}$ (muon and s-quark), and an underestimate for the particles that are above (charm, tau, (bottom ?)).

7.5.3 The π and K mesons

The π^0 mesons are bound states of the up and down quarks, that, differently from the proton and the neutron, "interact" with the $m_{3/10}$ scale through the electroweak coupling felt by their quarks, instead than directly through the strong force. As a consequence, the relation of the meson to the quark mass is given as according to 7.48, with α the electromagnetic coupling. We expect therefore:

$$\begin{aligned} m_{\pi}^2 \;&\sim\; \mathcal{O}(m_{\mathrm{q}}^2) \,\times\, \left\{ \alpha_{e.m.}\, \mathcal{O}(m_{3/10}^2) \,+\, \mathcal{O}(1) \right\} \\ &\approx\; \mathcal{O}(m_{\mathrm{q}}^2) \times \left\{ \alpha_{e.m.}\; (2m_{\mathrm{n}})^2 \,+\, \mathcal{O}(1) \right\}\,. \end{aligned} \tag{7.65}$$

This leads to a $\sim$ 100 MeV scale.

As we already observed, in principle the s-quark mass corrected by the $m_{3/10}$ scale as given in 7.62 is somehow already the effective mass "corresponding" to the K meson. It is not our scope here to enter into the details of the relation between the effective quark and meson mass, that, according to the common framework in which experimental data are interpreted, and therefore masses are derived, are supposed to be linked through $SU(3)$-colour-splitting relations. We want here only point out that, for what matters the charged mesons $\pi^{\pm}$ and $K^{\pm}$, they occupy a different phase space volume than the corresponding neutral ones; since the difference is due to the $U(1)_{\gamma}$ transformation properties, i.e. to the quark content, we expect the mass difference between charged and neutral mesons to be of the order of the mass difference of the component quarks. However, differently from the case of the neutron–proton mass difference, here we don't have stable particles. While for proton and neutron the phase space is basically the same (in they sense that they stably transform the one into the other), so that their differences simply reflect the differences in the properties of the bare

particles they are formed of, for the π and K mesons charged and neutral ones have access to completely different decay and interaction chains. Their phase spaces are therefore really different. As a consequence, although of the order of the mass difference of their quarks, the mass difference of the mesons are further modified by the modifications of the volumes of their effective phase spaces, and should be investigated as higher order corrections, after a recalculation of the phase spaces obtained by correcting the bare ones according to the meson interactions.

7.6 Gauge boson masses

We already discussed how, in any perturbative realization of a string vacuum with gauge bosons and matter states charged under a symmetry group, gauge and matter originate from T-dual sectors. For instance, in heterotic realizations the gauge bosons transforming in the adjoint of the group originate from the currents, while the fermions transforming in the fundamental representation originate from a twisted sector. As a consequence, any projection producing a non-vanishing mass for the matter states as the result of a shift on the windings, produces also a mass for the projected gauge bosons as a consequence of a shift on the momenta. Therefore, the mass of the projected matter states scales in a T-dual way to that of the gauge bosons. After the projections involved in the breaking of the internal symmetry into a set of separate families of particles, matter acquires a light mass, below the Planck scale, while the gauge bosons of the broken symmetry acquire a mass above the Planck scale.

As anticipated in section 3.3, the gauge bosons of the $SU(2)_{\rm w.i.}$ interaction don't follow this rule: they are lifted by a shift "on the momenta", not on the windings, and acquire an under-Planckian mass, of the same order as the matter states. As discussed in section 3.3, on the space-time coordinates act two shifts. One of them produces the breaking of parity, and gives a light mass to the matter states ($m \sim \mathcal{T}^{-1/2}$) while lifting in a T-dual way the mass of the gauge bosons coupled to the right-moving degrees of freedom. The other shift produces instead the lifting of the left-moving gauge group, the group of weak interactions. As discussed in section 3.3, this operation does not act, like the previous one, through "rank-reducing level-doubling", and its effect is not simply a further, equivalent shift. However, from the analysis of section 6.2.3, we see that more or less the effect of this ' 'Wilson-line like" operation is the one of producing a mass lift whose typical length is approximately a fourth root power, $\sim \mathcal{T}^{-1/4}$. The space-time gets in fact doubly contracted, the root being the lift up of what in a perturbative, logarithmic picture appear as projection coefficients, $\frac{1}{2}$ for the first shift and $\frac{1}{2} \times \frac{1}{2}$ from the second operation. Indeed, what we didn't say in section 3.3, is why parity should be broken by the first shift, and be related to the "square root" scale, and the weak group by the second shift, resulting approximately in a $\sim \mathcal{T}^{-1/4}$ scale [60]. Why not the opposite? In principle, there seems to be no ground for this choice of ordering. Once again, it is minimization of entropy what explains this choice: in the case of the first shift the phase space is reduced by the strength of the symmetry breaking driven by

[60] Indeed, as it can be seen from the mass expressions given in section 6.2.3, the highest mass scale is a $\sqrt{\alpha_{SU(2)}}^{-1} = \mathcal{T}^{1/56}$ factor above the the $\mathcal{T}^{-1/4}$ scale.

the mass of the right moving bosons, scaling as a positive power of the age of the Universe. The higher this power, the higher the reduction of entropy ($1/2 > 1/4$). In the case of the second shift, the boson masses scale as negative powers of the age of the Universe, and therefore the higher reduction of entropy, obtained with a higher boson mass, is achieved with a lower exponent ($1/4 < 1/2$).

The scale at which first the breaking of the $SU(2)_{\rm w.i.}$ symmetry takes place approximately corresponds the scale of the top-bottom mass difference, and, like the latter is higher than the experimental hadron mass scale, this one is around one-two orders of magnitude above the experimental mass scale of the $SU(2)_{\rm w.i.}$ bosons. It is reasonable to think that this one too is subjected to the same kind of renormalization as the other scales which are above the $m_{3/10}$ scale. However, as it is the case of the other masses, here too thinking in terms of shifts and elementary orbifold projections is a too simplified picture to get the fine details of mass differencies; in order to understand the mass of these bosons it is convenient to follow an approach similar to the one we have used for the masses of elementary particles, and use in our formulae the mass values already corrected according to section 7.5.2.

Let's first concentrate on what happens to the W bosons. The computation of the mass of these bosons proceeds similarly to that of the elementary particles. The basic diagram for a broken symmetry is again 6.42 of section 6.2.2, where, in this case, m_1 and m_2 refer to the masses of the heaviest $SU(2)_{\rm w.i.}$ doublet, the top-bottom quark pair. However, this time we are not interested in evaluating the ratio of the volumes occupied in the phase space by the up and down particle of the doublet. The relation between the two particles we want to consider is therefore not a $SU(2) = SU(2)_{\Delta m}$ symmetry, but the one established through the $SU(2)_{\rm w.i.}$ symmetry. Being produced by a shift "on the momenta", differently from the case of the broken symmetries discussed in section 6.2.2, here the W mass is T-dual to the mass of the over-Planckian bosons . Therefore, instead of 6.44, here the cut-off energy of the space of the process is:

$$< p >_{\rm w.i.} \;\sim\; \left(\frac{m_t m_b}{M_W^2} \right)^{1/4} , \tag{7.66}$$

and the relation 6.46 is replaced by:

$$\alpha \, \frac{3 m_t m_b}{M_W^2} \;\approx\; 1 \, , \tag{7.67}$$

where $\alpha \equiv \alpha_{SU(2)_{\rm w.i.}}$ The factor 3 can be understood in this way: each $SU(2)_{\rm w.i.}$ transformation rotates one quark colour; we need therefore three such rotations in order to pass from the bottom to the top quark. Notice that the relation 7.67 can be viewed as the integral form of a renormalization group equation. Differentiated and mapped to a logarithmic (and therefore also supersymmetric) representation, it roughly corresponds to the usual expressions of the beta-function:

$$\alpha \, \frac{m_t \, m_b}{M_W^2} \;\approx\; 1 \quad \stackrel{\partial, \log}{\rightsquigarrow} \quad b \;\approx\; T(R) - C(G) \, , \tag{7.68}$$

where b is the gauge beta-function coefficient and $T(R)$, $C(G)$ are the contributions of matter and gauge, entering with opposite sign. Inserting the mass values obtained in section 7.1.4,

as corrected in section 7.3, namely 7.32 and 7.33, and the value of the weak coupling 6.87, run at the bottom scale, $\alpha^{-1} \sim 24,1$ [61], we get:

$$M_{W^{\pm}} \sim 83,4\,\mathrm{GeV}\,. \tag{7.69}$$

In order to obtain this mass, we used for the top and bottom mass the "bare" values of page 135, not the values after the correction that brings them to their actual experimental value. Indeed, the relation 7.67 involves in its "bare" formulation bare particles. As it was for the quarks, also W are unstable and their mass is corrected by their interaction with the $m_{3/10}$ scale. However, for gauge bosons things go differently than for matter states, and their corrected mass cannot simply be obtained by plugging in 7.67 the corrected values of m_t and m_b. Gauge bosons behave T-dually with respect to particles; therefore, in their case, we must use an expression like 7.48 in its T-dual form:

$$\frac{1}{M_W^2} \to \frac{1}{M_W^2}\left(1 + \alpha \times \frac{1}{M_W^2}\int \frac{d^4p}{(p+m_{3/10})^2}\right), \tag{7.70}$$

where $m_{3/10}$ here basically stays for the neutron's mass, and the integral is intended up to the W-boson energy. Since $M_W > m_{3/10}$, $1/M_W < 1/m_{3/10}$, and, as in section 7.5.2, we correct the lower (inverse) scale $1/M_W$ with the higher (inverse) scale $1/m_{3/10}$. Moreover, the effective W-boson contact interaction is not suppressed by W-boson transfer propagators, and the strongest interaction they have with the $m_{3/10}$ scale occurs through the weak coupling. Therefore, here $\alpha = \alpha_{\mathrm{w.i.}}$. Owing to the different type of effective loop correction to the boson interaction with matter, as compared to the one of matter with matter, the term that multiplies the coupling is of order 1. We have therefore:

$$\frac{1}{M_W^2} \to \frac{1}{M_W^2}\,(1 + \alpha_{\mathrm{w.i.}})\,, \tag{7.71}$$

or, T-dualized back:

$$M_W^2 \to\approx M_W^2\,(1 - \alpha_{\mathrm{w.i.}})\,. \tag{7.72}$$

Inserting the value of $\alpha_{\mathrm{w.i.}}$ at the W mass scale, $\alpha^{-1}_{\mathrm{w.i.}}|_{M_W} \sim 23,46$, we obtain:

$$M_{W^{\pm}} \to \approx 81,6\,\mathrm{GeV}\,. \tag{7.73}$$

All the above expressions, 7.70, 7.71 and 7.72, neglect terms of order $\mathcal{O}(\alpha^2)$.

[61] In principle, also the weak coupling should undergo an effective beta-function modification similar to the one of the electromagnetic coupling discussed in section 7.4. However, as discussed in section 7.4 and in the footnote at page 142, this is expected to be a local modification, that tends to vanish toward the upper end scale of the matter sector, the scale that at present time is around the TeV scale. At the W-boson scale, $\alpha_{\mathrm{w.i.}}$ should have almost regained its "regular" value. However, we cannot exclude a slight modification toward a lower effective value, which could explain why we get a boson mass slightly higher than the experimental one. If we assume a "linear" decrease of the effect, from the MeV to the TeV scale, we should find that, if at the MeV scale the weak coupling undergoes a shift proportional to the one of the effective electromagnetic coupling: $\alpha_{\mathrm{w.i.}}|_{1\,\mathrm{MeV}} \to \alpha_{\mathrm{w.i.}}|_{1\,\mathrm{MeV}} \times (132,8/137)$, at the 80 GeV scale it should have lost $\sim 2/3$ of its effect, leading to a $\sim 83,0\,\mathrm{GeV}$ W-boson mass (see Appendix D).

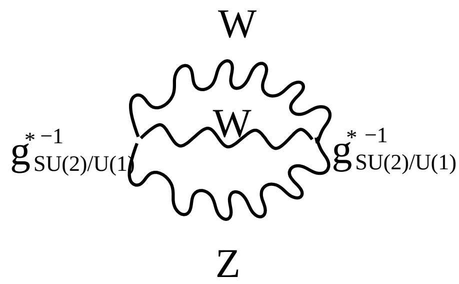
The mass of the Z boson cannot be directly derived in a similar way, by simply substituting m_t to m_b in 7.67: when $m_1 = m_2$ the symmetry is not broken, and the boson is massless! In first approximation, we expect the Z mass to be of the order of the mass of the W bosons: at the $SU(2)_{\rm w.i.}$ extended symmetry point the shift lifts all the three bosons by the same amount. However, as discussed in section 3.3, at the minimal entropy point, a bit away from the orbifold point, the group is broken to $U(1) = U(1)_Z$ through a parity-preserving, left-right symmetric operation. This means that what distinguishes the mass of the Z boson from the one of the chiral $W^{\pm}$ bosons is the fact that it acquires a "right moving" component: while the charged bosons interact only with a left-handed chiral current, the neutral boson has now a certain amount of coupling with a right-moving current [62]. This "misbalance" should be related to the mass difference between the Z and the $W^{\pm}$ bosons. In turn, as it is for all the other massive excitations, also the Z mass should be related to the volume it occupies in the phase space. The disagreement between the W and the Z mass should then be tuned by the strength of $SU(2)_{\rm w.i.}$ as compared to $U(1)_Z$. In order to derive the mass of the Z boson, consider therefore once again the diagram 6.42, this time with Z, W^- and W^+ replacing respectively the top, bottom quarks and the W boson: in this case we view the process as a transition between W^- and Z, produced by an element of the "group" $SU(2)_{\rm w.i.}/U(1)_Z$ (more precisely not a group but a coset) [63]. At the vertices, g is now the "coupling" of $SU(2)_{\rm w.i.}/U(1)_Z$. More precisely, since, as we discussed in section 6.2.5, the relation between "width" in the phase space and mass, in the case of gauge bosons, is the inverse with respect to the case of matter states (higher probability = lower boson mass), the relation of diagram 6.42 has to be "T-dualized" in the space of couplings; namely, "S-dualized". The coupling appearing at the vertices is therefore the inverse of the "coupling" g^* of $SU(2)_{\rm w.i.}/U(1)_Z$. This on the other hand is precisely what we should expect. If we set:

$$\alpha_{SU(2)_{\rm w.i.}} = \alpha^*_{SU(2)_{\rm w.i.}/U(1)_Z} \times \alpha_{U(1)_Z} \,, \tag{7.74}$$

being the $U(1)_Z$ coupling smaller than the one of the unbroken group, we obtain that $\alpha^* > 1$, and the relation 6.42 must be dualized in order to reduce to the ordinary weak coupling diagram:

[62] This is not a general property of string vacua: when going to the $U(1)$ point, it is in general not true that vector fields get mixed up. In the present case, these fields are massive, already lifted by a shift, after which the left and right chiral components of fermions combine to give rise to massive matter. From a "pure" string point of view the $SU(2)$ bosons don't exist as massless fields, and therefore not as gauge bosons. Here we are discussing about the properties of massive string excitations, that we account among the field degrees of freedom only because their mass is small, lower than the Planck scale. Since the parity-breaking shift reflects in a "level-2" realization of the weak group, there is no surprise that a displacement toward the $U(1)$ point indeed involves the breaking of two $U(1)$'s, the left and right, as a matter of fact "patched together".

[63] Notice that we are not defining here Z as a linear combination of W^0 and the field B_μ, associated to the hypercharge.



(7.75)

Instead of a momentum to the fourth power, now the loop integral pops out a momentum squared, and at the place of 6.44, the typical momentum of the space of the process is here:

$$< p > \ \sim \ \frac{M_Z}{M_W} \, . \tag{7.76}$$

The W mass, the mass of the boson mediating the process, appears in the denominator, as in 7.66, T-dually to the case of 6.44. From the diagram 7.75 we obtain therefore:

$$\left(\frac{M_Z}{M_W} \right)^2 \ \approx \ \alpha^{*}_{SU(2)_{\text{w.i.}}/U(1)_Z} \ , \tag{7.77}$$

and, using the relation 7.74,

$$M_Z \ \sim \ \sqrt{\frac{\alpha_{SU(2)_{\text{w.i.}}}}{\alpha_{U(1)_Z}}} \, M_W \, . \tag{7.78}$$

In order to obtain $\alpha_{U(1)_Z}$ we can proceed as in section 6.2.5, this time by determining the fraction with respect to the volume occupied by $SU(2)_{\text{w.i.}}$ at the place of $SU(2)_{\Delta m}$. This means that the coupling of $U(1)_Z$ should stay to the coupling of $U(1)_\gamma$ in the same ratio as the coupling of $SU(2)_{\text{w.i.}}$ stays to the one of $SU(2)_{\Delta m}$. Therefore, we expect:

$$\frac{\alpha_{U(1)_Z}}{\alpha_{SU(2)_{\text{w.i.}}}} \approx \frac{\alpha_{U(1)_\gamma}}{\alpha_{SU(2)_{\Delta m}}} \, . \tag{7.79}$$

At present time, 6.81 and 6.85 and the W-boson mass 7.73 tell us that the Z boson mass should be approximately:

$$M_Z \ \sim \ 1,127 \, M_W \quad \approx 91,96 \, \text{GeV} \, . \tag{7.80}$$

If we proceed as in the footnote at page 148, by assuming a linear decrease of the local correction to the effective beta-function, this time of the electromagnetic coupling discussed in section 7.4, till its vanishing at the top scale of the charged matter phase space, 7.18, we get that at the 80 GeV scale the shift should have been reduced to around $1/4$ of its size, producing a relative modification of the electromagnetic coupling at this scale of a factor $\sim 1,00791$, leading to a modification of the ratio 7.78 by a factor $1,00394553$, i.e. a Z to $W^{\pm}$ mass ratio:

$$\frac{M_Z}{M_W} \ : \ 1,127 \ \to \ 1,132 \ , \tag{7.81}$$

a number that should be compared with the experimental ratio of these masses, $\sim 1,134$ [58]. Owing to the theoretical uncertainties implicit in our derivation, it does not make sense to refine the calculation, although it seems that the linear approximation of the effective beta-function is not quite far from the real behaviour.

Let's now come back to see how, for what concerns the neutral and charged currents, the low-energy action looks like. According to the results of section 3.3, the matter states feel:

i) a long range force, mediated by a massless field, with the strength of the non-anomalous traceless $U(1)$ group discussed in section 6.2.5,

ii) a short range force, mediated by the bosons obtained from the breaking of $SU(2)_{(\mathrm{L})}$ ($= SU(2)_{\mathrm{w.i.}}$), and

iii) a would-be "very short" range force, "mediated" by the boson corresponding to T_3^{R}. As we said, this third is not an interaction in the sense of field theory, because the mass of this boson is higher than the Planck mass. On the other hand, its contribution is highly suppressed.

Our analysis tells us that $SU(2)$ *singlets, both left and right moving*, couple to a massless boson in a diagonal way with the strength of the group $U(1)_{\tilde{Y}}$, whose present time value is given by 6.85. The electric charge corresponds to a certain choice of charge distribution among the degrees of freedom constituting these singlets, as dictated by minimization of entropy. It can be viewed as a linear combination of the three $U(1)$ charges T_3^{L}, T_3^{R} and $\tilde{Y}$, resulting in a traceless "shift" in the definition of the states. The total charge remains of course the same as the "hypercharge" $\tilde{Y}$, and therefore also the strength of the coupling, which is related to the "width" of the interaction [64]. We conclude therefore that this is the strength of the electromagnetic interaction, mediated by a massless boson that we identify with the photon.

The left-handed current couples in non-diagonal way to the W^+ and W^- bosons associated to the "raising" and "lowering" generators of the $SU(2)_{\mathrm{w.i.}}$ group, and diagonally with the $U(1)_Z$ resulting after the breaking of this group. Owing to the fact that this symmetry is broken, the coupling and the mass of this boson is not the same as the one of the $W^{\pm}$ bosons: they are related as given in 7.78. If we now consider the values of the couplings of these terms, "run back" to an infinite-volume effective action in order to make possible a comparison with the standard description of low-energy physics, we can see that approximately α_γ, $\alpha_{\mathrm{w.i.}}$ and "α_Z", the total coupling of the Z boson, are related by ratios that can be written as:

$$\sqrt{\alpha_\gamma} \approx \sqrt{\alpha_{\mathrm{w.i.}}}\,\sin\theta\,; \tag{7.82}$$

$$\sqrt{\alpha_Z} \approx \sqrt{\alpha_{\mathrm{w.i.}}}\,\cos\theta\,, \tag{7.83}$$

[64] This in first approximation: mass differences lead in fact to a differentiation of the decay widths, depending on the mass/charge assignments and distribution.

for a certain angle θ. The value of α_γ, the coupling of the electromagnetic current $J^{e.m.}$ to the photon, numerically approximately coincides with the one of the fine structure constant. $\alpha_{\rm w.i.}$ is the coupling of the $W^\pm$ and Z bosons to the axial current $J^{\pm,0}$, and numerically approximately coincides with the standard weak coupling α_w. Notice that in our framework also the Z boson couples to this current by the same amount as the charged bosons. In the Standard Model effective action, at the tree level the couplings of the $J^{\pm,0}$ are instead:

$$W^\pm \;\to\; g\, J^\pm_\mu W^{\mp\,\mu}\,, \tag{7.84}$$

$$Z \;\to\; \frac{g}{\cos\vartheta_w} J^0_\mu Z^\mu\,. \tag{7.85}$$

However, the effective amplitude is weighted also by the boson mass, in such a way that:

$$\alpha_{\rm eff}(W) \;\sim\; \frac{\alpha_w}{M^2_W}\,, \tag{7.86}$$

$$\alpha_{\rm eff}(Z) \;\sim\; \frac{\alpha_w}{\cos^2\vartheta_w}\times\frac{1}{M^2_Z} \;=\; \frac{\alpha_w}{\cos^2\vartheta_w}\times\frac{\cos^2\vartheta_w}{M^2_W} \;=\; \frac{\alpha_w}{M^2_W}\,. \tag{7.87}$$

Effectively, the two couplings are therefore the same. This agrees with the fact that, in our framework, couplings are related to the effective interaction/decay width, not to the microscopical description in terms of gauge strengths and connections. In the Standard Model, the Z boson couples also to the electromagnetic current, in such a way that its total interaction with matter is:

$$\frac{g}{\cos\vartheta_w}\left(J^0_\mu \,-\, \sin^2\vartheta_w J^{e.m.}_\mu\right)\, Z^\mu\,. \tag{7.88}$$

This is a consequence of the fact that the boson mass eigenstates are obtained from the hypercharge and the Cartan of the $SU(2)_{\rm (L)}$ group through an orthogonal transformation. The "total" width of the Z boson is therefore corrected by the fact that it couples also to a left-right symmetric current, whose trace is the same as the one of our "hypercharge". Effectively, the size of the coupling is:

$$\alpha^{\rm eff}_Z \;\sim\; \frac{1}{4\pi}\frac{g^2}{\cos^2\vartheta_w}(1-\sin^2\vartheta_w)^2 \;=\; \alpha_w\cos^2\vartheta_w\,. \tag{7.89}$$

In practice, this means that, for what concerns the axial coupling, the Z boson behaves analogously to the charged bosons, namely, it couples with the same strength and it "propagates" with the same mass. However, it is also "displaced" by a left-right symmetrically interacting term, in such a way that its total width corresponds to a higher mass and lower coupling, as given in 7.89. These relations should be compared with 7.82 and 7.83. The empiric parametrization we gave in terms of the angle θ corresponds then to the expressions in terms of the Weinberg angle ϑ_w:

$$\theta \;\cong\; \vartheta_w\,. \tag{7.90}$$

Therefore, for what concerns the electromagnetic and the charged and neutral axial currents, we have effectively reproduced, although in a completely different way, the coupling terms

of the effective action of the Standard Model. Things turn out to correspond, although the underlying basic description doesn't at all. On the other hand, perhaps it is not so dramatic that things don't exactly match with all the details of the Standard Model description. It should be clear that our way of approaching all the issues related to the low energy parameters is completely different from the approach of the Standard Model, or its field theoretical extensions, even "string inspired" ones. In our case, there is no field theoretical Higgs mechanism at work; masses are generated, and symmetries are broken, in a different way, and consequently different is also the parametrization of the elementary interactions. Our aim is not to show how in detail a Standard-Model-like description of the "microscopical physics" is recovered. The usual effective theory is anyway an approximation, a "choice of linearization", which, although if appropriately restricted to a certain region of the space of parameters it can nicely fit the experimental data, on a larger scale it seems to not work. Certainly, fitting of data shows a rather good agreement of many Standard Model predictions with the experiments. However, this agreement is achieved by advocating, "predicting", the existence of degrees of freedom that continue to escape any attempt of detection. We can almost perfectly fit the Weinberg angle and boson masses, but at the expense of introducing a Higgs field with a mass such that it should have been already detected somewhere, and still some things don't match. We can further adjust the misprediction by enlarging the number of degrees of freedom, which allow to "improve the direction", pushing forward the problem. For instance, detecting supersymmetric partners at low energy etc. In this situation, what is the meaning of statements like "the Standard Model correctly predicts the one or the other quantity", if at the end this is the result of a fit which assumes the existence of non-detected degrees of freedom? Of course, these questions are well known since a long time, and the answer still open. We recall the problem here just to remind the reader that there is nothing illegal in the fact that we are not reproducing all the terms of the low-energy Standard Model action, apart from those aspects that one can safely consider as "experimentally detected", and not just "fitted".

A remark on the meaning of "massless" in this framework

In the usual field theoretical representation of particles and interactions in infinitely extended space-time, a free photon can not be "localized" as can a particle, and its energy can be as small as we want. In our framework, this corresponds only to an asymptotic situation, closer and closer approached but concretely never realized. At present time, a photon can not be more extended than the distance from us to the horizon of the Universe. Masslessness does not translate therefore in the energy beeing arbitrarily small, but just in the property that "the minimal energy corresponds to the inverse of the radius of the Universe":

$$E_\gamma \sim \frac{1}{2}\frac{1}{\mathcal{T}} \,. \tag{7.91}$$

Massive particles and fields are characterized by a minimal energy that departs from 7.91 in that it scales as a lower power of the inverse of the Universe, and therefore they have an extension smaller than its size:

$$E_{\text{massive}} \sim \frac{1}{2}\frac{1}{\mathcal{T}^p} \,, \qquad 0 < \, p \, < 1 \,, \tag{7.92}$$

so that the spread in space is:

$$< \Delta X >_{\text{massive}} \ \sim \ \mathcal{T}^p \ < \ \mathcal{T} \, . \tag{7.93}$$

For a massive object, the exponent p is necessarily always smaller than 1.

7.7 The Fermi coupling constant

We are now in a position to make contact with the experimental value of the weak coupling. This is measured through the so-called Fermi coupling constant G_F, a dimensional ($[m^{-2}]$) parameter defined as the effective coupling of the weak interaction at low transferred momentum [65]:

$$\frac{G_F}{\sqrt{2}} \ = \ \frac{g^2}{8M_W^2} \ = \ \frac{\pi \alpha_w}{2M_W^2} \, . \tag{7.94}$$

From section 7.6 we know that we can identify $\alpha_{\text{w.i.}}$ with the usual weak coupling α_w of the literature Inserting our results for the W-boson mass, 7.73, and the value of the weak coupling at the W-boson scale, given at page 148, we obtain:

$$G_F|_{M_W} \ = \ 1,4221 \times 10^{-5} \, \text{GeV}^{-2} \, . \tag{7.95}$$

As it was for the case of the fine structure constant, once again we are faced with the problem of understanding what is the meaning of a physical quantity, whose value is always related to a certain experimental process at a certain scale. Here, from an experimental point of view the Fermi coupling is obtained by inspecting the pion into muon decay. The effective, infinite-volume renormalization of G_F to the pion–muon scale is obtained in our framework in the same way as for the other couplings, namely treating G_F as a generic coupling, whose behaviour is represented through an effective linearization as in 6.94. The relative variation from the W to the $\mu \div \pi$ scale [66] is of order:

$$\frac{\Delta G_F}{G_F}|_{M_W \to m_\pi} \ \approx \ 0,81 \, , \tag{7.96}$$

and we get:

$$G_F|_{\pi/\mu} \ \approx \ 1,1519 \times 10^{-5} \, \text{GeV}^{-2} \, , \tag{7.97}$$

a value about 1 % away from the effective experimental value [58]. The percent is on the other hand the order of the precision we have in our estimate of the W-boson mass, and as a consequence we cannot hope to get something better for the Fermi coupling.

⋆ ⋆ ⋆

This brief survey does not pretend to exhaust the argument of masses. In particular, certainly not the topic of meson and baryon masses. However, we hope to have at least shed a bit of light onto the subject. Within the scenario we are discussing, the "entropy

[65] Low means here negligible when compared to the W-boson mass.

[66] Within our degree of approximation, it does not make such a difference the choice of one scale or the other, between muon and pion.

approach" seems to provide a powerful tool for the investigation of both perturbative and non-perturbative masses.

As we mentioned in section 6.2.11, for the practical purpose of computing the fine corrections to masses and couplings it may turn out convenient to map to an appropriate "linearized representation" of the string space, in which the issue of the computation of these quantities with the methods of geometric probability can be approached with a more evolved technology. What however is missing in these approaches is a "cosmological perspective", which would give the possibility of "fixing the scale". For instance, the agreement of (almost) all the quantities computed in Ref. [64] with experiments is impressive, and, according to what we discussed in section 6 (in particular, 6.2.11), in many cases it is not just mere coincidence or numerology; however, at least one input has to be supplied from outside; e.g. the electron's mass, or any other scale, to serve as the "measure" to which to compare all the ratios of volumes. Moreover, we expect that in order to refine the results such as those of [64], [66] or [67], a better understanding of how the non-perturbative vacuum is mapped, and its degrees of freedom are represented in such a linearized approximation, is in order.

8 Mixing flavours

As is well known, mass eigenstates are not weak-interaction eigenstates: the weak currents cross off-diagonally the elementary particles, and, besides diagonal up/down decays, there is a smaller fraction of non-diagonal, flavour changing decays that mix up the three families. Experimentally, this phenomenon is well known for what matter quarks. Still hypothetical is however whether it occurs also for leptons: its detection would go in pair with a clear indication of non-vanishing neutrino masses. Indeed, in this case one tries to go the other way around, looking for neutrino oscillations as a signal of non-vanishing neutrino masses. In the case of quarks, the standard parametrization of these phenomena is made through the introduction of a matrix $V_{\rm CKM}$, the Cabibbo-Kobayashi-Maskawa matrix, which encodes all the information about the "non-diagonal" propagation of elementary particles. It is defined as the matrix which rotates the base of "down" quarks of the $SU(2)$ doublets, allowing to express the currents eigenstates in terms of mass eigenstates:

$$V_{\rm CKM} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} . \tag{8.1}$$

$V_{\rm CKM}$ accounts for the mixing among different generations, as well as the CP violations.

8.1 The effective CKM matrix

In our framework, masses have a different explanation than in ordinary field theory. Nevertheless, it is still possible to refer to an effective field theory description, once accepted that this has to be considered only as a tool useful for practical purpose, without any pretence of being a (self-)consistent theory. Although in our approach we directly consider decay amplitudes rather than effective terms of a Lagrangian, in order to make possible an easy comparison between our predictions and the usual literature it is therefore convenient building an effective Lagrangian, that, within the rules of field theory, will reproduce, or "mimic", our amplitudes.

As is well known, the CKM matrix is a unitary matrix, in which all phases except one are reabsorbed into a redefinition of the quarks wave functions. Therefore, its nine entries are parametrized by nine real coefficients and one phase, responsible for parity violation. In our framework, the analysis of the spectrum has been carried out by classifying the degrees of freedom according to their charge. This means that what we got are the "current eigenstates" (section 3). Subsequently, we have considered a "perturbation" of this configuration, obtained by switching-on so-far neglected degrees of freedom, in order to investigate their masses (section 6). In section 6.2.2 we have put mass ratios in relation with ratios of sub-volumes of the phase space, which is divided into several sectors by the breaking of the initial symmetry. Mass ratios are then related to the couplings of the broken symmetries. As we anticipated, there are two kinds of breaking: a "strong breaking", in which the would-be gauge bosons acquire a mass above the Planck scale, and a "soft breaking", in which the gauge bosons acquire a mass below the Planck scale. Only in this second case the transition

appears as an ordinary decay, mediated by a propagating massive boson. Otherwise, the boson of the broken symmetry works somehow like an external field: we don't see any boson propagating, and we interpret the phenomenon as a "family mixing". The off-diagonal entries of the CKM matrix precisely collect the effect of this type of "non-field-theory decay": off-diagonal entries account for transitions from one generation to another, non mediated through gauge bosons as in the case of ordinary decays.

According to our previous discussion, the ratios between entries of the CKM matrix should be of the same order of the mass ratios, normalized to the full decay amplitude: mass ratios correspond in fact to squares of "elementary" couplings: $m_f/m_i \sim \alpha_{i\to f}$ $(\sim g^2_{i\to f})$. If α_{ab} is the coupling for the flip from family a to family b, the decay amplitude of a $a \to b$ flavour changing decay is expected to be proportional to α^2_{ab}.

Any decay amplitude depends on masses, both of initial and final states of the process. With our non-perturbative methods we have direct access to the full decay amplitudes. In order to make contact with the ordinary description of the mixing mechanism, we must consider that, as it is defined, the CKM matrix is unitary, and collects the information about flavour changing, subtracted from any dependence on masses: in expressions of amplitudes, this dependence is carried by other terms. This allows to normalize the matrix in such a way that, owing to the fact that off-diagonal elements are much smaller than diagonal ones, the diagonal elements are close to 1:

$$|V_{ud}|,\ |V_{cs}|,\ |V_{tb}| \approx 1\,, \tag{8.2}$$

and, with a good approximation,

$$|V_{us}| \approx |V_{cd}| \tag{8.3}$$
$$|V_{ub}| \approx |V_{td}| \tag{8.4}$$
$$|V_{cb}| \approx |V_{ts}| \tag{8.5}$$

As for the computation of masses, a detailed evaluation of the CKM matrix entries would require taking into account all processes contributing to the determination of the phase space. Here we want just to make a test of reliability of our scheme; we are therefore only interested in a first approximation. To this purpose, it is reasonable to work within the framework of the simplifications 8.2–8.5. Owing to these simplifications, we can restrict our discussion to the off-diagonal elements $|V_{ts}|$, $|V_{td}|$ and $|V_{cd}|$. A direct, non-diagonal $t \to s$ decay should have an amplitude of order m_s/m_t, normalized then through m_b/m_t in order to reduce to the scheme 8.2. A rough prediction for V_{ts} is therefore:

$$V_{ts} \approx \frac{m_s}{m_b} \sim \frac{0,147\,\mathrm{GeV}}{3,6\,\mathrm{GeV}} \sim 0,04\,, \tag{8.6}$$

where we have used the values 7.61 and 7.54. Similarly, we obtain:

$$V_{td} \approx \frac{m_d}{m_b} \sim 0,001\,, \tag{8.7}$$

and

$$V_{cd} \approx \frac{m_d}{m_s} \sim 0,027\,. \tag{8.8}$$

While 8.6 basically agrees with the commonly expected value of this entry (see Ref. [58]), 8.7 and 8.8 are away by a factor ~ 4 in the first case, and ~ 8 in the second. An adjustment of the value is not a matter of "second order" corrections. Here the problem is that for these mixings, experimental results are mostly obtained through branching ratios of meson (π, K) decays. In these quark compounds, the strong non-perturbative resummation is highly sensitive to the GeV scale. Indeed, an experimental value $|V_{cd}| \sim 0,22$ seems to be much influenced not by the mass ratio of the bare quarks, but of the K and π mesons:

$$V_{cd} \approx \frac{m_\pi}{m_K} \sim \mathcal{O}(0,22)\,. \tag{8.9}$$

Although in a lighter way, the meson scale seems to modify also the ratio of the bottom to down quark transition. As we already said, here it is not a matter of determining a physical quantity: only decay amplitudes are physical, the CKM matrix doesn't have a physical meaning in itself. It is therefore crucial to see how do we refer to this effective tool: how much "resummation" we want to attribute to a correction to be applied to "bare" decay amplitudes computed from a "bare" CKM matrix, and how much of it we prefer to already include in the CKM matrix. As long as the final products we consider are just meson amplitudes, the two approaches are equivalent.

By comparing eqs. 8.8 and 8.9, we are faced with something at the same time reasonable and which nevertheless sounds somehow odd. On one hand, the fact that 8.9 gives a higher ratio is not surprising: it is in fact quite natural to think that a heavier particle has a larger decay probability than a lighter one. On the other hand, when applied to the $|V_{cd}|$ transition, this argument seems to lead to a contradiction: the basic degrees of freedom of a K- and π-mesons are the quarks; nevertheless, the Kaon has a larger decay probability than the quarks it is made of. Indeed, it is not in this way that the enhancement of the V_{cd} entry due to the passage from quarks to mesons has to be interpreted. The free quark "does not exist", Pions and Kaons are the lightest strong-interaction singlets containing the d and s quark. Once inserted in the computation of a decay amplitude, the values we are proposing for the entries of the CKM matrix must be corrected by some overall "form factor", of the order of m_K/m_s for the initial state, and of m_π/m_d for the final state. In practice, this is equivalent to the introduction of an "effective" CKM matrix entry, $V_{cd}^{\text{eff}} \sim (m_K/m_s)/(m_\pi/m_d) \times V_{cd}$. This rescaling eats the factor ~ 8 of disagreement between our prediction and the usual value of this entry, as reported in the literature.

Differently from the case of $|V_{cd}|$, $|V_{ts}|$ turns out to be in agreement with what reported in the literature, because the latter is derived by unitarity from $|V_{cb}|$, measured through $B \to D$ decays. Both these mesons have a mass of the same order as the b and c quark respectively. To be more precise, in these cases the quark mass itself, as is given in the literature, corresponds to the "corrected mass", basically coinciding with the mass of the meson of which it constitutes the heaviest component. The matrix entry is therefore "by definition" almost the same as the "bare" one.

The case of the $|V_{td}|$ (and V_{ub}) entries is even more involved, being much higher the uncertainties in the experimental derivation of the transition elements.

A more detailed derivation of the CKM entries as a function of masses can be found in Ref. [64]. As we said, this is perfectly fine in a neighbourhood of our present time. In our case,

the entries of 8.1 are however time-dependent, and the branching ratios of various decays vary along the evolution of the Universe. The various particles tend toward a higher relative separation: although the curvature of space-time tends to a "flat" limit, and the absolute value of masses decrease with time, the ratios of the various masses increase, thereby lowering the probability of mixing among families. This goes together with the T-dual increase with time of the mass of the "would be gauge bosons" of the broken symmetry among generations, the non-field-theoretical decay we discussed in section 6.2.2.

In our framework, neutrinos are massive, and we expect that the CKM matrix has a leptonic counterpart. The "leptonic CKM" entries should however be more suppressed, as a consequence of the rearrangement of the phase space, so that all the three neutrinos are lighter than the lightest charged lepton, and their spaces have a higher separation. According to the leptonic mass values derived in section 7.1.1, we expect at present time approximately:

$$
\begin{aligned}
V_{\mathrm{CKM}}^{\mathrm{leptons}} & \equiv \begin{pmatrix} V_{e\nu_e} & V_{e\nu_\mu} & V_{e\nu_\tau} \\ V_{\mu\nu_e} & V_{\mu\nu_\mu} & V_{\mu\nu_\tau} \\ V_{\tau\nu_e} & V_{\tau\nu_\mu} & V_{\tau\nu_\tau} \end{pmatrix} \\
& \approx \begin{pmatrix} \sim 1 & \sim 0,007 & \sim 0,00005 \\ \sim 0,007 & \sim 1 & \sim 0,007 \\ \sim 0,00005 & \sim 0,007 & \sim 1 \end{pmatrix} .
\end{aligned} \tag{8.10}
$$

Non-diagonal lepton decays are therefore more difficult to observe than those of quarks, perhaps more difficult to detect than neutrino masses themselves.

8.2 CP violations

In our framework, CP (and T) violation is already implemented in the construction: it is a consequence of the shifts in the space and time coordinates, that break parity (in the case of space), and time reversal symmetries. As we have seen, this is related to entropy, and to the fact that also the second principle of thermodynamics is automatically implemented in this scenario.

Let's consider the decay of a particle, e.g. $K \to \pi\,(+\gamma)$, or $B \to K\,(+\gamma)$. The decay probability is proportional to the square "coupling": $\alpha^2_{K\to\pi\gamma}$, $\alpha^2_{B\to K\gamma}$. The square effective couplings represent volumes in the phase space: they are in fact proportional to mass ratios, and can be viewed as the inverse of a proper time, the proper time of the decaying particle measured in units of the decay product, raised to the fourth power. The larger is this volume, the higher is the decay probability.

However, if we consider the problem more in detail, we see that not the entire length, given as the inverse of the mass, is at disposal for increasing the entropy through the decay process: part of this phase space is occupied by the rest energy of the product particle(s), that we collectively indicate with m_f. As for all masses, the origin of m_f is a shift along space-time. We want now to see what happens if we invert the shifted coordinate. Let's consider the simplified case of just one space-time coordinate, t. In this case, this inversion

is a time reversion. Under $t \to -t$, the shift operation turns out to act in the opposite direction, and results in an expansion of the volume. In our framework, time is compact: $t \in [0, \mathcal{T}]$. A mass corresponds to a shift $t \to t + a$, $a \approx \frac{1}{m_f}$, such that now the zero point has been displaced to a: $0 \to a$. If we perform a time reversal operation, $t \to -t$, we obtain that the shift acts now as: $(-t) \to (-t) + a$. By overall sign reversal, this is equivalent to a mirror situation, $t \to t - a$, in which we have a "particle" with mass $-m_f$. Roughly speaking, we can think that the volume in the phase space occupied by this particle is "stolen" from the correct time evolution, and goes "in the opposite direction". The asymmetry between a "straight" decay and its "time-reversed" one is given by the fact that in the first case the phase space volume is proportional to $[(m_i + m_f)/m_f]^4$, after the inversion to $[(m_i - m_f)/m_f]^4$. The general result with four space-time coordinates is then a simple consequence of the fact that a global inversion of all the space-time coordinates: $t \to -t, \vec{x} \to -\vec{x}$ is a symmetry of the system.

In decays involving transitions from neighbouring generations. e.g. $b \to s$, $s \to d$, the mass ratio is at present time $m_i/m_f \sim 10$ and $\sim 3,8$ respectively (as in the case of the CKM angles, the mass ratios we have to consider are not those of free quarks, but of the particles effectively involved in the decays. In this specific case, the B, K and π mesons). Therefore, at present we get an asymmetry of order $\sim (1/3,8)^4 \approx 4,8 \times 10^{-3}$ for the K decays, and $\sim 10^{-4}$ for B decays.

In the case of D mesons ($(c\bar{d})$, $(c\bar{u})$ and conjugates), we get a $\mathcal{O}(10^{-2})$ decay asymmetry. This D result is of particular relevance, because in this case there is no reliable prediction based on the Standard Model effective parametrization. Indeed, while in our framework we obtain an asymmetry in the range of the experimental observations, the Standard Model prediction fails to account for the magnitude of the observed effect (for a review, see for instance [58]). Because of this, CP violations in the D mesons decays are often considered a test for "new physics".

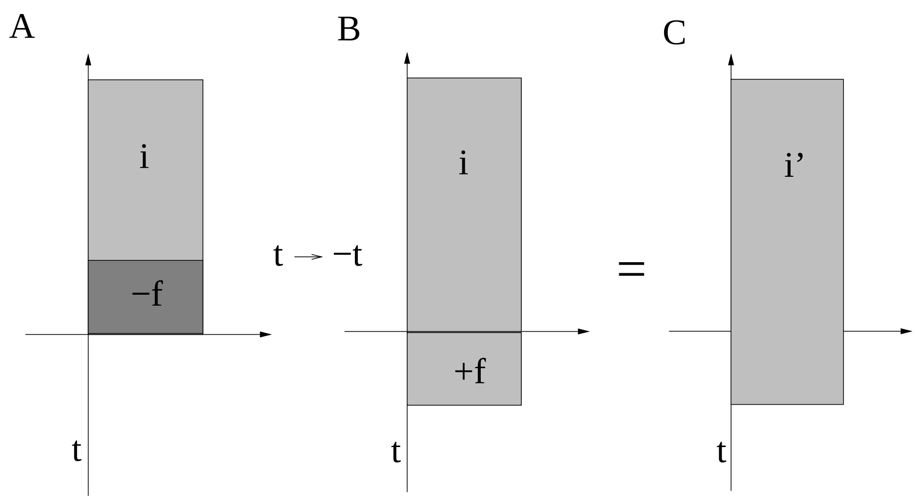


Figure 11: The increase in the decay amplitude as produced by a time reversal. While in picture A the phase space volume of the final state f has to be subtracted from the volume at disposal of the initial state, after a time flip it adds up (picture B), ending in an increased decay probability for the initial state (picture C).

9 Astrophysical implications

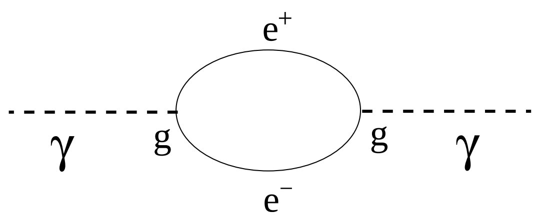
Expression 2.31 contains in principle all the information about the Universe, at any time of its evolution. In section 5 we have already seen how it is possible to derive information also about "astronomical" data, such as the cosmological constant and the energy density of the Universe. We have then investigated the masses and interactions of the elementary particles. Masses and couplings are of interest for the physics of accelerators. According to the analysis of our scenario, we don't expect breaking news from high energy colliders, apart from sectors such as the neutrino physics and CP violation parameters. A larger potential source of stringent tests seems to come from astrophysics, a field which appears the more and more as one of the most exciting domains of present and future investigation. This is particularly true for our framework, in which the present day physics is tightly related to the history and the evolution of the Universe. Therefore, after having gained a better insight into the details of the elementary particles, here and in section 10 we come back to consider some important issues addressed by astrophysical observations.

9.1 The CMB radiation

An important experimental cosmological observation is the detection of a "ground" cosmic electromagnetic radiation with the typical spectrum of a black body radiation, with a temperature of about 2,8 ^0K [69, 70]. This phenomenon is often claimed to constitute a proof of the theory of the Big Bang: this radiation would consist of photons cooled down during the expansion of the Universe, and at the origin they should have possessed an energy corresponding to a microwave length, as expected from energy exchange due to Compton scattering through the plasma at the origin of the Universe. Here we want to discuss how these issues are addressed in our scenario. In the framework we are proposing, the history of the Universe is already implied in the solution of 2.31; its expansion and cooling down are already "embedded" in the framework, which is in itself a cosmological scenario, and in principle don't need to be added as separate inputs. As a consequence, the 3 Kelvin temperature of this radiation can be justified by directly using the present-day data of the Universe. This does not mean that the CMB radiation does not have any relation with the Universe at early times of its expansion: simply, the present configuration of the Universe is in our framework a particular case of a cosmological solution accounting for any era of its evolution, and therefore already "contains the information" about the earlier times. In particular, all energies and mass scales of our present time are the primordial ones, cooled down.

According to the discussion of section 3.3, the spectrum of the Universe "stabilizes" into neutrons/protons+electrons (the visible matter) plus free neutrinos ν_e, photons and, of course, gravitons. The CMB radiation consists of photons of "ground energy", i.e. they are not produced by interactions and decays of visible matter. This background radiation can be viewed as a bath of photons in contact with a "thermal reservoir", a neutral vacuum, whose energy corresponds to the inverse square root scale of the age of the Universe, the ground

energy scale of matter [67]. Particle's pair production out of this vacuum leads eventually to photon production. In order to understand what the average energy of this "photon sea", the "temperature", is, let's look at a typical photon/vacuum transition process, taking into account that at equilibrium the photon propagator, intended as "the mean propagator", is not renormalized. As we already discussed several times, in the physical picture Feynman-like diagram corrections act multiplicatively, a passage from logarithms of quantities such as couplings and coordinates to their exponential being subtended. The traditional series expansion corresponds to a linearization, to working with algebras, on the "tangent space", of a phenomenon that, once resummed, has to be viewed as working at the level of the group. The diagram representing the class of processes of interest for us is sketched in figure 9.1.



(9.1)

As seen from the "exponential picture", diagrams like this one contribute to a multiplicative correction to the photon propagator, such that:

$$\frac{1}{p_\gamma^2} \to \frac{1}{p_\gamma^2} \times \left[\frac{\alpha \int (\text{loop})}{p_\gamma^2} \right], \tag{9.2}$$

where $\alpha = \alpha_{\text{e.m.}} \equiv \alpha_\gamma$ and the loop integral pops out a $[m^2]$, corresponding to the energy of our black body reservoir E_0:

$$\int (\text{loop}) \approx E_0^2 \approx \left(\frac{1}{2\sqrt{\mathcal{T}}} \right)^2. \tag{9.3}$$

In other words, we can say that the ground electromagnetic radiation consist of a gas of photons at equilibrium, interacting with a matter vacuum, whose ground energy fluctuations are, according to the Uncertainty Principle, of the order of 1/2 the inverse of the typical length of the matter sector, the square root scale of the age of the Universe. Since at equilibrium, as we have said, the average photon propagator is not renormalized, we conclude that the term under square brackets in 9.2 must contribute by a factor 1, and therefore:

$$< p_\gamma > \sim \sqrt{\alpha_\gamma} \, \frac{1}{2} \, \frac{1}{\sqrt{\mathcal{T}}}. \tag{9.4}$$

[67]Notice that we must use the ground energy scale, not the "mean" mass scale, roughly corresponding to the neutron mass scale. Here we are not interested in the non-perturbative mass eigenstate at any finite time, but in the ground energy step of a Universe evolving toward a "flat" limit at infinity (flat in the sense of the non-compact orbifolds discussed in the previous sections, in particular in section 5.3).

Using the value of α_γ at the electron's scale, derived through an effective running of the type 6.94 from the initial value 6.85 at the $\mathcal{T}^{-1/2}$ scale to the $0,5\,\mathrm{MeV}$ scale, $\alpha_\gamma^{-1}|_{m_e} \sim 132,3$, 6.19 for the present age of the Universe, and converting energy into temperature through the Boltzmann constant, we obtain:

$$T_\gamma \equiv k^{-1} < p_\gamma > \; = k^{-1} E_\gamma^0 \sim 2,72\,^0 K \,. \tag{9.5}$$

A signal that the present day phenomenon results from the cooling down of a primordial situation comes from the spatial inhomogeneity of this radiation over a solid angle of observation [71]. We have seen that in our framework, that describes the physics "on shell", the invariance under space rotations is broken by the same mechanism that gives rise to non vanishing masses for the matter states. The scaling of masses, that, going backwards in time, tend to the Planck scale, tells us that in the primordial Universe also the space inhomogeneity must have been higher than what it is today. Toward the Planck scale, the spatial inhomogeneities are expected to become of a size comparable with the size of the Universe itself.

9.2 The fate of dark matter and the Chandra observations

A discrepancy between our framework and the common expectations is the absence in our scenario of dark matter. According to our analysis, the Universe consists only of the already known and detected particles. Of course, there can be regions of the space in which a high concentration of neutrinos, which for us are massive, increases the curvature without being electromagnetically detected. But this is not going to change dramatically the scenario: there is no hidden matter acting as an extra source able to increase the gravitational force by around a factor ten over what is produced by visible matter, as it seems to be required in order to explain a gravitational attraction among galaxies much higher than expected on the base of the estimated mass of the visible stars. The problem arises in several contexts: Big-Bang nucleosynthesis, rotational speed of galaxies, gravitational lensing. All these points would require a detailed investigation, beyond the scope of this work. We will also not attempt to rediscuss a huge literature, and limit ourselves here to mention some hypotheses. The first remark is that the discrepancies between theoretical expectations and the observed effects, which are found in so different issues as primordial Universe, nucleosynthesis and galaxy phenomenology, don't need necessarily to be explained all in the same way.

About nucleosynthesis we will spend some words in section 10.3. Let's consider here the problems related to the motion of external stars in spiral galaxies, where for the first time the issue of dark matter has been addressed, and the "anomalous" gravitational lensing, with reference to the recently observed effect in the 1E0657-558 cluster [72].

It is since 1933 (Fritz Zwicky) that, by looking at the amount of red-shift in the light emitted by the stars in the wings of a spiral galaxy, it has been noticed how, differently from what expected, the rotation speed does not decrease with the inverse of the square root of the radius: it is a constant [73, 74]. Presence of invisible matter has been advocated, in order to fill the gap between the mass of the observed matter and the amount necessary to increase the gravitational force. Indeed, the expectation that the rotation speed of stars in

the external legs should decrease is based on the assumption that almost the entire mass of the galaxy is concentrated in the bulge at the center of the spiral. Any star on the wings would therefore feel the typical gravitational field due to a fixed, central mass.

In the framework of our scenario, masses have been in the past higher than what they are now. Moreover, owing to the fact that, as we discussed in section 2.1, the Universe "closes up", in such a way that the horizon we observe corresponds to a "point", the space separation between objects located at a certain cosmic distance from us appears to be larger than what actually is. All this could mean that the mass of the center of a galaxy, as compared to the wings, has been systematically overestimated. It would be interesting to see, by carrying out a detailed re-examination of the astronomical observations, whether the behaviour of the center of a galaxy still requires to advocate the presence of a heavy black hole, in order to explain a gravitational force higher than what expected on the base of the estimated mass of the visible stars. In any case, it is possible that, once the downscaling of length and upscaling of masses has been appropriately taken into account, a better approximation of a spiral galaxy is the one sketched in figure 12. In part A of the picture the galaxy is (very roughly) represented with wide wings, with stars relatively "broadened" on the plane of the galaxy. Part B shows the same figure, simply with much narrower arms. In picture A the broad lines have been shadowed in a way to make evident that the higher star density of the bulge is largely due to the "superposition" of the various arms. Nevertheless, as it is clear from picture B, the problem remains basically "one-dimensional": the wings are one-dimensional lines coming out of the center of the galaxy. Under the hypothesis that all the stars have the same mass, the linear density of a wing is constant, and, once integrated from the center up to a certain radius R, the total mass M_R of the portion of galaxy enclosed within a distance R from the center is roughly proportional to R:

$$\rho = \frac{dM}{dr} \sim \text{const.} \quad \Rightarrow \; M_R \sim \text{const} \times R \, . \tag{9.6}$$

In the expression of the gravitational potential, the linear R dependence of the mass cancels against the R appearing in the denominator (the potential remains the one of a Coulomb force). The gravitational potential energy is therefore a constant times the mass of the star in the wing. Conservation of energy implies therefore that also the velocity of the star does not depend on the radius R. We stress that this is only an approximation: it would be exact if the arms were not those of a spiral but straight legs coming out radially from the center, and under the assumption that all the stars of the bulge correspond to the superposition of the arms.

In the case of the 1E0657-558 cluster, the Chandra observatory has detected a gravitational lensing higher than what expected on the base of the amount of luminous matter. Moreover, the highest effect corresponds to two dark regions close to the cluster, rather than to places where the visible matter is more dense. In the framework of our scenario, a possible explanation could be that what is observed is the effect of a "solitonic" gravitational wave, produced as a consequence of the separation of one sub-cluster from the other one. This could increase the gravitational force by an amount equivalent to the displaced cluster mass, for a length/time comparable to the cluster size, therefore a time much higher than the few hours during which the effect has been measured ($\sim$ 140 hours). It remains

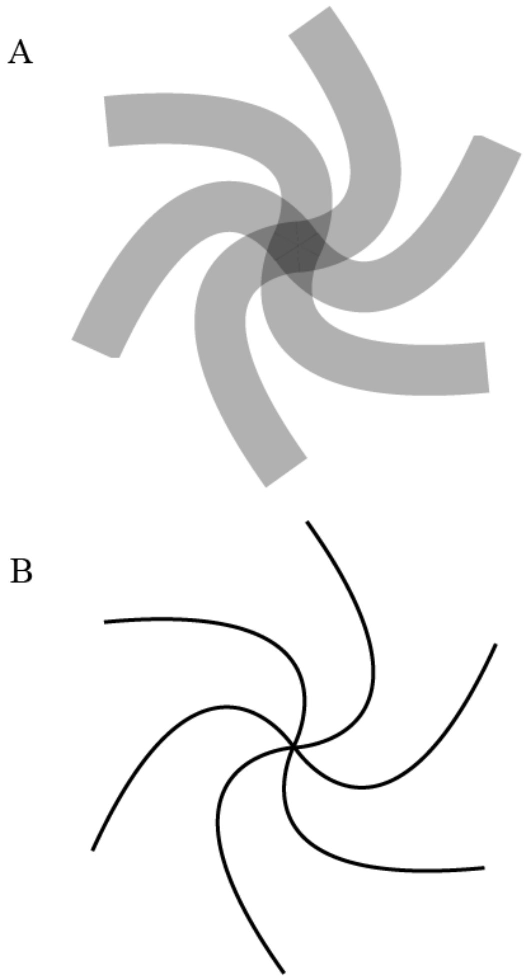


Figure 12: Picture A is the rough sketch of a spiral galaxy, in which the arms are broad and shadowed in a way to highlight the increasing mass density due to their superposition at the center. Figure B represents the same object, with the arms narrowed down, in order to highlight the one-dimensional nature of the physical problem, for what concerns the mass density.

that the lensing is around 8-9 times higher than what expected on the base of the amount of visible mass. However, the cluster under consideration is at about 4 billion light years away from us. This is around 1/3 of the age of the Universe. This time distance is large enough to make relevant the effects due to a change of the curvature of space-time along the evolution of the Universe, as well as a change of masses, according to 5.8–5.10 and 6.10. Furthermore, as we discussed above, the apparent space separation between objects located at a certain cosmic distance from us must be appropriately downscaled, in order to account for the curving up of space-time into a sphere, with the horizon "identified" with the origin. Putting all this together, we obtain that the effective gravitational force experienced on the 1E0657-558 cluster is (or, better, it was) indeed 8-9 times higher than what it appears to us on the base of the expected mass of the objects in the cluster, i.e. precisely the amount otherwise referred to dark matter.

10 Cosmological constraints

Recently, cosmology has addressed two kinds of problems for what concerns the "running back" of a theory, or an "early time" model. Namely, 1) the possible non-constancy of what are commonly called "constants", and 2) the agreement with the expected origin/evolution of the early Universe (baryogenesis, nucleosynthesis etc...). In our framework, these issues are put in a light quite different from the usual perspective: there are in fact indeed no constants; therefore, a variation of couplings, masses, cosmological parameters, and, as a consequence, energy spectra, is naturally implemented. However, there is a peculiarity: all these parameters scale as appropriate powers of the age of the Universe. As a consequence, a "number" close to one at present day has a very mild time dependence:

$$\mathcal{O}(1) \approx \mathcal{T}^{\epsilon} \Rightarrow |\epsilon| \ll 1\,, \tag{10.1}$$

and therefore varies quite a little with time. Oklo and nucleosynthesis bounds, being given as ratios of masses and couplings that cancel each other to an almost "adimensional" quantity, are precisely of this kind. In our case they don't provide therefore any dangerous constraint.

For what concerns the non-constancy of "constants", there are not enough data enabling to test our prediction about a time variation of the cosmological constant, whose measurement is still too imprecise. A more stringent test of the variation of parameters comes from the observations on the light emitted by ancient Quasars. In this case, the spectrum shows an "anomalous" red-shifted spectrum. This shift should not be confused with the usual red-shift, of which we have discussed in section 5.2. The effect we consider here persists once the "universal" red-shift effect has been subtracted. As an explanation, it is often advocated a possible time variation of the fine structure constant α. We already devoted a paper to this subject [3], at a time in which many issues concerning our framework were not enough clear. We therefore rediscuss here the argument, in the light of a better understanding of the theoretical framework we are proposing.

10.1 The "time dependence of α"

The question of the possible time variation of the fine structure "constant" arises in the framework of string theory derived effective models for cosmology and elementary particles. Various investigations have considered the possibility of producing some evidence of this variation, or at least a bound on its size. To this regard, astrophysics is certainly a favoured field of research, in that it naturally provides us with data about earlier ages of the Universe. A possible signal for such a time variation could be an observed deviation in the absorption spectra of ancient Quasars [75, 76, 77]. This effect consists is a deviation in the energies corresponding to some electron transitions, which remains after subtraction of the background effect of the red-shift, and is obtained with interpolations and fitting of data.

What is observed is a decrease of the relativistic effects in the energies of the electrons cloud, with respect to what expected on the base of present-day parameters (in particular, the fine structure constant). Indeed, while the atomic spectra are universally proportional to the atomic unit $me^2 \propto m\alpha^2$, the relativistic corrections depend on the coupling α. After

subtraction of the "universal" red-shift effects, their variation should then be directly related to a variation of α. As an explanation of this effect, in Ref. [3] we proposed the increase of the electron's mass, when measured with respect to the typical matter scale, the scale to which in our scenario the size of the red-shift is related. At that time we didn't know the time dependence of α, although we already guessed that, going backwards in time, it had to increase, rather than decreasing, as it would be required in order to justify a decrease of the relativistic effects.

By now we know that indeed α scales as a positive power of the inverse of the age of the Universe, and that it goes back to one at the Planck scale. Apart from the specific type of scaling (power instead of logarithm), this behaviour agrees with what one would normally expect, namely that the coupling tends backwards to a unification value, where it meets the weak and strong couplings. After a reconsideration of the problem with a better knowledge of the behaviour of masses and couplings, we conclude that the argument we proposed in [3] is not convincing; in our framework, the explanation comes from considering both the scaling of α and the one of masses at the same time. It is in fact true that, going backwards in time, α increases, as also the proton and the electron mass do. However, *the ratio of α to the mass scales decreases*. This is the "variation of α after subtracting the universal red-shift" which is usually considered in the discussions of the literature. Namely, if we measure the variation of α with respect to the electron's mass scale (whether the true electron mass or the "reduced" mass doesn't make a relevant difference [68]), i.e. if we rescale quantities in the frame in which masses are considered fixed, we indeed observe a decrease of the coupling α. Indeed, what is done in the literature (see Refs. [77]) is not only to consider masses fixed, but to exclude from the evaluation also the effect of the red-shift. With current experimental methods, based on the interpolation of spectral data in order to find out the "background" and the variations out of it, this subtraction is somewhat unavoidable.

In order to obtain what the prediction in our scenario is, and how it compares with the literature, let's first see how the decrease of the relativistic effects, when going backwards in time, turns out to be a prediction of our framework. Consider the effective scaling of α in terms of $m\alpha^2$ units, the "universal" scaling of emission/absorption atomic energies. We have that:

$$\bar{\alpha} \stackrel{\text{def}}{\equiv} \frac{\alpha}{m\alpha^2} \approx \mathcal{T}^{\frac{1}{3}+\frac{1}{28}} . \tag{10.2}$$

The "effective" coupling $\bar{\alpha}$ scales as a positive power of the age of the Universe: going backwards in time, it decreases. According to the literature, atomic energies have an approximate scaling of the type [69]:

$$E_n \approx K_n (m\,\alpha^2) + \Gamma_n \, \alpha^2 (m\,\alpha^2) , \tag{10.3}$$

where K_n and Γ_n are constants and the second term, of order α^2 with respect to the first one, accounts for the relativistic corrections. Investigations on the possible variation of α use interpolation methods to disentangle the second term from the first one. Since the universal part is reabsorbed into the red-shift, the relative variation should give information on just

[68] In the hydrogen atom this is given by $m_e = \frac{m_e m_{\rm p}}{m_e + m_{\rm p}}$. The possibility of referring to a change of this quantity the effect measured in Ref. [77] can be found in Ref. [78, 79, 80].

[69] See for instance Ref. [77]

the variation of α. Expression 10.3 is of the form:

$$E_n \approx E_n^0 (1 + a_1 \alpha^2) . \tag{10.4}$$

It is derived by considering the first terms of a field theory expansion around the fine structure constant (the electric coupling). Indeed, since we are interested in the correction subtracted of the universal part reabsorbable in the red shift, we can separate the $\mathcal{O}(\alpha^2)$ term in 10.3 as:

$$K_n = (K_n - \Gamma_n) + \Gamma_n . \tag{10.5}$$

This allows to reduce the part of interest for us to:

$$E_n^{\text{eff}} \approx E_n^0 (1 + \alpha^2) . \tag{10.6}$$

As we already observed several times along this work, perturbative expressions involving elementary particles are naturally defined and carried out in a logarithmic representation of the physical vacuum. In particular, when writing expressions like 10.3 it is intended that the coupling α scales logarithmically. An expression like 10.6 should be better viewed as accounting for the first terms of a series that sums up to an expression scaling as a certain power of the age of the Universe:

$$\alpha_{\text{eff}} \equiv 1 + \alpha^2 \approx 1 + \alpha^2 + \mathcal{O}(\alpha^4) \rightsquigarrow \exp \alpha \sim \mathcal{T}^{\beta} , \tag{10.7}$$

where α is then not the full coupling, intended in the non-perturbative sense of 6.51, but its logarithm. According to 10.2, in the hypothesis of keeping masses fixed, this term should then effectively scale as a positive power of the age of the Universe: $\beta > 0$. The exponent β can be fixed by comparing values at present time:

$$\alpha|_{\text{today}} \approx \sqrt{5 \times 10^{-5}} . \tag{10.8}$$

We obtain therefore:

$$1 + \alpha^2 \approx \mathcal{O}(1 + 5 \times 10^{-5}) \approx \mathcal{T}^{\beta} \Rightarrow \beta \sim \mathcal{O}(10^{-6}) , \tag{10.9}$$

and a relative time variation:

$$\frac{\dot{\alpha}_{\text{eff}}}{\alpha_{\text{eff}}} \approx \beta \mathcal{T}^{-1} \approx \mathcal{O}\left(10^{-16} \, \text{yr}^{-1}\right) . \tag{10.10}$$

This is the relative variation of the relativistic correction subtracted of the universal part (reabsorbed in the red-shift), to be compared with the results of [77]. Since the deviation of the resummed function 10.6 from a pure exponential is of order $\alpha^4 \sim 2 \times 10^{-9}$, four orders of magnitude smaller than the dominant term, the inaccuracy in our computation is much lower than the order of magnitude of the result.

10.2 The Oklo bound

Data from the natural fission reactor, active in Oklo around two billions years ago, are today considered one of the most important sources of constraints on the time variation of the fundamental constants. By comparing the cross section for the neutron capture by Samarium at present time with the one estimated at the time of the reactor's activity, one derives a bound on the possible variation of the fine structure constant, and on the ratio $G_F m_{\rm p}^2$, in the corresponding time interval. The interpretation of the experimental measurements and their translation into a bound on the variation of the capture energy resonance is not so straightforward, and depends on several hypotheses. In any case, all these steps are sufficiently under control. More uncertain is the translation of this bound on the energy variation into a bound on the variation of the fine structure constant and other parameters: this passage requires strong assumptions about what is going to contribute to the atomic energies. This analysis was carried out in Ref. [81], basically on the hypothesis that the main contribution to the resonance energy comes from the Coulomb potential of the electric interaction among the various protons of which the nucleus of Samarium consists. According to [81], after a certain amount of reasonable approximations, the energy bound translates into a bound on the variation of the electromagnetic coupling. A simple look at expression 6.80 shows that, in our scenario, the variation of this coupling over the time interval under consideration violates the Oklo bound. This bound seems therefore to rule out our theoretical framework. However, things are not so simple: the derivation of a bound on a coupling out of a bound on energies works much differently in our framework, and we cannot simply use for our purpose the results of [81]. Indeed, in our framework what varies with time is not only the fine structure constant, but also the nuclear force, and the proton and neutron mass as well. Of relevance for us is therefore not a bound on a coupling, derived under the hypothesis of keeping everything else fixed, but the bound on the energy itself [81]:

$$-\,0,12\,\mathrm{eV} \;<\; \Delta E \;<\; 0,09\,\mathrm{eV}\,. \tag{10.11}$$

In order to give an estimate of the amount of the energy variation over time, as expected in our framework, we don't need to know the details of the evaluation of the resonance energy starting from the fundamental parameters of the theory. To this purpose, it is enough to consider that, whatever the expression of this energy is, it must be built out of i) masses, ii) couplings (electro-weak and strong) and iii) the true fundamental constants (the speed of light c, the Planck constant $\hbar$, and the Planck mass $\mathrm{M_p}$). Working in units in which the latter are set to 1 (reduced Planck units), all parameters of points i) and ii) scale as a certain power of the age of the Universe. As a consequence, the resonance energy itself mainly scales as a power of the age of the Universe:

$$E \;\sim\; a\mathcal{T}^{-b}\,. \tag{10.12}$$

(More generically, it could be a polynomial: $E \sim a_1\mathcal{T}^{-b_1} + a_2\mathcal{T}^{-b_2} + \ldots + a_n\mathcal{T}^{-b_n}$. In this case, to the purpose of checking the agreement with a bound, it is enough to look at the dominant term). We can fix the exponent b by comparing the expression, evaluated using the present-day age of the Universe, with the value of the resonance, that we take from [81]:

$$E \;\sim\; a\mathcal{T}^{-b} \;=\; 0,0973\,\mathrm{eV} \;\times\; 1,2\,\times\,10^{-28} \;=\; 1,2\,\times\,10^{-29}\,\mathrm{M_P}\,. \tag{10.13}$$

In order to solve the equation, we would need to know the coefficient a, something we don't. However, as long as we are just interested in a rough estimate, it is reasonable to assume that, since this coefficient mostly accounts for possible symmetry factors, it may affect the value of the result for about no more than one order of magnitude. Inserting the value $\mathcal{T} \sim 5 \times 10^{60} \mathrm{M}_{\mathrm{P}}^{-1}$ for the age of the Universe, we obtain:

$$b \sim \frac{1}{2}, \tag{10.14}$$

and finally:

$$|\Delta E| \sim \frac{1}{10} E \sim 0,01 \text{ eV}. \tag{10.15}$$

over a time of two billion years. This is compatible with the Oklo bound, eq. 10.11.

From the Oklo data one tries also to derive a bound on the adimensional quantity

$$\beta \equiv G_F m_{\mathrm{p}}^2 (c/\bar{h}^3). \tag{10.16}$$

In this case, our discussion is easier, because we know the scaling of all the quantities involved [70]. Once again, we have to deal with a quantity that scales as a power of the age of the Universe. At present time, we have:

$$\beta \sim \mathcal{T}^{-b_\beta} = 1,03 \times 10^{-5}. \tag{10.17}$$

Inserting the actual value of the age of the Universe, we obtain $b_\beta \sim \frac{1}{12}$. Over a time interval of around 1/5 of the age of the Universe, this gives a relative variation:

$$\frac{\Delta \beta}{\beta} \sim 0,017, \tag{10.18}$$

to be compared with the one quoted in Ref. [81]:

$$\frac{|\beta^{\mathrm{Oklo}} - \beta^{\mathrm{now}}|}{\beta} < 0,02. \tag{10.19}$$

Both results 10.15 and 10.18, although still within the allowed range of values, seem to be quite close to the threshold, beyond which the model is ruled out. One would therefore think that a slight refinement on the measurement and derivation of these bounds could in a near future decide whether it is still acceptable or definitely ruled out. Things are not like that. Indeed, as we already stressed in several similar cases, the *entire* derivation of bounds and constraints, involving at any level various assumptions about the history of the Universe and therefore of its fundamental parameters, should be rediscussed within the new theoretical framework: it doesn't make much sense to compare pieces of an argument, extracted from an analysis carried out in a different theoretical framework, with different phenomenological

[70] We recall that $G_F/\sqrt{2} = g^2/8M_W^2$. Therefore, $\beta = \pi \alpha m_{\mathrm{p}}^2/\sqrt{2}M_W^2$. For times much higher than 1 in reduced Planck units, the proton mass can be assumed to scale approximately like the mean mass scale 6.10.

implications. To be explicit, in the case of the derivation of the Oklo bounds, one should reconsider all the derivation of absorption thresholds and resonances. We should therefore better take into account from the beginning the time variation of all masses, and in particular the neutron and proton masses, as well as couplings. Perhaps a more meaningful quantity is then not anymore the pure resonance shift, but this quantity rescaled by the neutron mass. In this case, the effective variation of interest for our test is not 10.15, but:

$$\frac{\Delta(E/m_n)}{(E/m_n)} \approx \frac{\Delta \mathcal{T}^{-\frac{1}{9}}}{\mathcal{T}^{-\frac{1}{9}}} \sim 0,02\,, \tag{10.20}$$

a variation one order of magnitude smaller than 10.15 ($\Delta E/E \sim 0,1$). Analogous considerations apply also to the case of the second bound 10.18, basically equivalent to the nucleosynthesis bound.

10.3 The nucleosynthesis bound

Bounds derived from nucleosynthesis models are even more questionable: they certainly make sense within a certain cosmological model, but, precisely because of that, they cannot be simply translated into a framework implying a rather different cosmological scenario. Once again, the only anchor points on which we can rely are the few "pure" experimental observations, to be interpreted in a consistent way in the light of a different theory. The point of nucleosynthesis is that there is a very narrow "window" of favourable conditions under which, out of the initial hot plasma, our Universe, with the known matter content, has been formed. Of interest for us is the very stringent condition about the temperature (and age of the Universe) at which the amount of neutrons in baryonic matter have been fixed. As soon as, owing to a cooling down of the temperature, the weak interactions are no more at equilibrium, the probability for a proton to transform into a neutron is suppressed with respect to the probability of a neutron to decay into a proton. Owing to their short life time, comparable with the age of the Universe at which the equilibrium is broken, basically almost all neutrons rapidly decay into protons, except for those that bound into ^{4}He. Nucleosynthesis predicts a fraction of ^{4}Helium and Hydrogen baryon numbers ($\sim 1/4$) in the primordial Universe, which is in good agreement with experimental observations. The formula for the equilibrium of neutron/proton transitions is given by:

$$\frac{\mathrm{n}}{\mathrm{p}} = \mathrm{e}^{-\frac{\Delta m}{kT}} \sim 1\,, \tag{10.21}$$

where $\Delta m = m_{\mathrm{n}} - m_{\mathrm{p}}$. In the standard scenario, this mass difference is a constant, and the temperature runs as the inverse of the age of the Universe. The equilibrium is broken at a temperature of around $0,8$ MeV, when $(\mathrm{n/p}) \simeq 1/7$. In our framework too the temperature runs as the inverse of the age of the Universe, but the mass difference Δm is not a constant: all masses run with time. At large times ($\mathcal{T} \gg 1$ in Planck units), we are in a regime in which we can use the arguments of section 7.5, in order to conclude that, being the u and d quark masses much lighter than the neutron mass scale (which is related to the "$m_{11/36}$" mass scale), we can consider Δm as a perturbation of $m \simeq m_{\mathrm{n}}$. In this regime, the neutron-proton mass difference is basically of the order of the constituent quark mass difference,

and we have reasons to expect that it also runs accordingly. It would therefore seem that, in our case, going backwards in time, the ratio (n/p) remains lower than in the standard case, and the equilibrium 10.21 is attained at a temperature much higher. However, to the purpose of determining the processes of the nucleosynthesis, essential is not just the scaling of the equilibrium law of the neutron-to-proton ratio, but also that of the mean life of the neutron. It is the combined effect of these two quantities what determines the primordial baryon composition. In the standard case, the neutron mean life is assumed to be constant. Being related to the neutron decay amplitude, i.e. to the volume occupied by the neutron in the phase space, in our framework this quantity too is not constant. In order to see what in practice changes in our scenario with respect to the standard one, instead of attempting to guess what the scaling behaviour of the neutron mean life could be, we can proceed by considering, instead of the pure running of the equilibrium equation, the *reduced running at fixed neutron mean life*. Certainly the mean life is constant if the neutron mass is constant. The quantity of interest for us is therefore the scaling of the mass difference, as measured in units of the neutron mass itself. According to our considerations of above, we have:

$$\Delta m_{\mathrm{red}}(\mathcal{T}) \equiv \frac{\Delta m}{m_{\mathrm{n}}} \sim \frac{\mathcal{T}^{p_{(u-d)}}}{\mathcal{T}^{p_{\mathrm{n}}}}, \tag{10.22}$$

where $p_{(u-d)}$ and p_{n} are exponents corresponding to the up-down quark mass difference and to the neutron mass respectively. This running is expected to hold not only at present time but also at a temperature of ~ 1 MeV, which is anyway much lower than the Planck scale. We can therefore compare our prediction with the standard one by simply considering the relative deviation of equation 10.21 from its standard value, as obtained by replacing the constant mass difference Δm with $\Delta m_{\mathrm{red}}(\mathcal{T})$:

$$\frac{\mathrm{n}}{\mathrm{p}} = \mathrm{e}^{-\frac{\Delta m}{kT}} \quad \to \quad \left(\frac{\mathrm{n}}{\mathrm{p}}\right)_{\mathrm{red}} \equiv \mathrm{e}^{-\frac{\bar{m}_{\mathrm{n}} \Delta m_{\mathrm{red}}(\mathcal{T})}{kT}}, \tag{10.23}$$

where $\bar{m}_{\mathrm{n}}$ is the *fixed*, time-independent present-day value of the neutron mass. Therefore, in the standard case $(\mathrm{n/p})_{\mathrm{red}}$ coincides with (n/p). According to the mass values given in section 6, we have:

$$\Delta m_{\mathrm{red}}(\mathcal{T}) \approx \mathcal{T}^{-\frac{1}{24}}. \tag{10.24}$$

Considering that the time variation between the point $\mathcal{T}_f$ of the breaking of equilibrium and the present day is of the order of the age of the Universe itself, $\Delta T \equiv \mathcal{T} - \mathcal{T}_f \sim \mathcal{T}$, we obtain approximately that the integral variation of $x \equiv \Delta m_{\mathrm{red}}(\mathcal{T})$ over this time interval is:

$$\Delta x \sim \frac{1}{24} x. \tag{10.25}$$

The "reduced value" of (n/p), $(\mathrm{n/p})_{\mathrm{red}}$, is now modified to:

$$\left(\frac{\mathrm{n}}{\mathrm{p}}\right)_{\mathrm{red.}} : \frac{1}{7} \to \sim \frac{1}{7}\left(1 - \frac{\ln 7}{24}\right) \approx 0,131. \tag{10.26}$$

This value leads to a ratio X_4 of helium to Hydrogen of around:

$$X_4 \sim 0,232, \tag{10.27}$$

still in excellent agreement with what expect from today's most precise determinations (for a list of results and references, see Ref. [58]).

As mentioned above, there is here no reason to push the discussion into further detail, because the entire issue, as well as all the extrapolations from experimental observations, should be rediscussed within the framework of this scenario, something well beyond the scope of this work. We want however to point out another aspect of the problem, which arises in our theoretical framework. A peculiarity of our scenario is that, at any time, there is a bound on the number of particles that can exist in the Universe. At any time the volume of space-time is finite; the maximal amount of energy is fixed by the Schwarzschild Black Hole relation, $2M = R$, where R is here identified with $\mathcal{T}$, and the lowest particle's mass is the one of the electron's neutrino, given in 7.2. The number of matter states is therefore finite, and cannot exceed the Schwarzschild mass of the Universe divided by the lowest neutrino mass. On the other hand, this statement has the "mean value" validity of any other statement related to the class of configurations that dominate in 2.31. As we already discussed, only in an "average" sense we can in fact talk of geometry of space-time, and make contact with "classical" objects such as Black Holes, and in general with the Einstein's equations and their Scharzschild's solution. One may then ask if the "Black Hole bound" on the total mass of the Universe can be "violated" in a quantum way, for a time $\Delta t \sim 1/\Delta m$, also along the minimal entropy solutions: although they are the configurations with a space-time more close to a classical geometry description, they are nevertheless string vacua and, as such, quantum vacua. The point is however how do we observe such a violation: in order to measure a mass fluctuation $\Delta m \sim 1/\Delta t$, we need precisely a time $\Delta t > 1$ in Planck units. According to the Black Hole bound, in a time Δt the Universe increases its energy from $E \sim \mathcal{T}$ to $E' \sim \mathcal{T} + \Delta t$, i.e. $\Delta E \sim \Delta t$. Since $\Delta t > 1$, we have $\Delta E > \Delta m$. In other words, the mass/energy fluctuation implied by the Uncertainty Principle is always lower than the total energy increase due to the time evolution of the Universe: in the time we need in order to measure a possible quantum fluctuation out of the ground value of the energy, this value changes by a much larger amount. Quantum fluctuations are smaller than the "classical" increase due to the expansion [71].

[71] To be more precise, there is nothing classical here: we have seen that precisely the fact of being the interior of a black hole makes the Universe a quantum system.

11 Concluding remarks

In this work we investigated the consequences of the finiteness of the speed of light in a Universe bounded by a horizon of observation. Light is the "medium" that makes possible our perception of all what exists. It is therefore somehow obvious to expect that its properties affect our experimental measurements. However, this in principle could just constitute a secondary effect, to be "subtracted" in order to get the true, "intrinsic"nature of what happens in the Universe. The point of view of this work is instead much more radical: we have assumed that, since any information about all what we know to exist and we can experimentally measure comes to us at the speed of light, the "Universe" is at all the effects constituted by our causal region, defined as the set of all what we can "experimentally" observe. Notice that, by this, we intend to include also things that maybe have not yet been detected, but which are linked to us by a light line shorter than the horizon corresponding to the age of the Universe. But, more importantly, we exclude things which fall out of our causal region (our horizon), and whose existence can only be inferred within a specific theoretical framework. The only assumption is that such a horizon exists [72], and that the "visible part" of the universe is no more to be intended as a truncation, a subset of the whole that exists: it underlineis the whole that exists. This interpretation has dramatic consequences on the geometry of space-time.

Under these conditions, all what we observe turns out to be a consequence of the finiteness of the speed of light. Actually, light is not the only quantity travelling at speed c: there is also gravity, and our perception of the Universe is mediated also by the gravitational force. Indeed, similarly to what happens in the case of the electric and magnetic fields, also light and gravity, i.e. electromagnetism and gravitation, turn out to be dual aspects of the same phenomenon.

We have seen how the existence of a horizon corresponding to the "big bang" point is a crucial input that, through the theory of Relativity, implies a non-vanishing curvature, and therefore a non-vanishing energy, of such a causal region, "the Universe". The size of space-time Δt and the minimal energy fluctuation of the "vacuum" ΔE turn out to be precisely related by the Heisenberg's uncertainty relation:

$$\Delta E \, \Delta t \geq \frac{1}{2} \, . \tag{11.1}$$

The Universe manifests therefore itself through a wave-like medium (light and/or gravity), and is endowed with an Uncertainty Principle; this implies that it possesses a quantum nature [73]. Further investigation reveals that the quantum gravity theory underlying its description must come with a built-in T-duality: the candidate is therefore String Theory. Indeed, the uncertainty relation 11.1 looks like some kind of T-duality relation: it seems to

[72] The existence of an horizon itself is a conclusion derived within a specific theoretical framework: what is actually observed is just an expansion of the Universe, from which one infers the existence of a starting point. The existence of a horizon is then a consequence of the finiteness of the speed of light; however, in itself running back in time an equation is not really an experiment.

[73] See Ref. [1] for a discussion of the subtleties related to the probabilistic interpretation of dynamics usually associated to Quantum Mechanics.

mean that at times/lengths above the Planck scale ($\Delta t > 1$) energy fluctuations are below the Planck scale, and vice-versa. Therefore, time and energy would be T-dual with respect to the Planck scale. This is not the T-duality we usually encounter in string theory, which relates a radius to its inverse, and energies with momentum number to energies with winding number:

$$E \approx \frac{n}{t} \ \rightarrow \ n\, t \,. \tag{11.2}$$

Indeed, as a "macroscopic" relation, 11.1 characterizes a situation of *broken T-duality*: it is the statement of a *covariance*, what remains after the breaking of the T-duality invariance. In the case of unbroken T-duality, the space of "times" and the space of "energies" are equivalent, so that we can say that unbroken T-duality maps energies to energies: the generic string energy expansion contains both the momentum and winding terms. Our world, i.e. 11.1, corresponds to a situation in which only momentum energies, i.e. Kaluza-Klein modes of the type $E \sim m/R$, with $R > 1$, are observed; in a compact space-time, talking about effective action of light degrees freedom makes only sense if T-duality is broken. When T-duality is broken, the space of energies becomes equivalent to the T-dual of the space of times. Otherwise, over- and sub-Planckian worlds would be equivalent and there would be no "inequality" 11.1: the inequality is a signal of broken symmetry. After the breaking of T-duality, a space/time inversion maps therefore a space/time coordinate to its conjugate in the classical sense, i.e. momentum/energy. This operation implements therefore the wave-particle duality.

The breaking of T-duality, as well as the entire dynamics of the Universe, including the fact that space-time extends only along four coordinates, do not result as the solution of a particular equation: if we consider all possible string configurations, it turns out that the configuration of our world simply correspond to the class of configuration which are more often realized in the phase space. Namely, we can explain all what we observe by simply assuming that the world is a superposition of string configurations; there are many more string configurations in which T-duality is broken than configurations with un-broken T-duality, and the world therefore "looks more like" a solution with broken T-duality. Similarly it goes for the dimensionality of extended space-time, the breaking of supersymmetry, and in general the appearance of the low-energy world. The entropy principle itself, as well as any kind of C-theorem, becomes here just an "average statement". All this is encoded in expression 2.31, namely:

$$\mathcal{Z} \ = \ \int \mathcal{D}\psi \, \mathrm{e}^{-S} \,, \tag{11.3}$$

where the exponential of S, the entropy, precisely expresses the fact that the "weight" of a configuration is given by the product of the probabilities of the states, elevated to themselves. This can be considered as some kind of "quantization of quantization": the Heisenberg's Uncertainty Principle itself, and therefore the quantum nature of our world, are "mean value statements", that work for the "dominant" configuration of our world. As much as in quantum (field) theory we have a weighted sum over classical trajectories, here we have a weighted sum over string vacua, i.e. over "quantum field configurations". The "dynamics" implied by 11.3 says that the system "evolves" mostly through minimal increase of S, i.e. preferably to the nearest, most similar configurations. There is however

no "solution" in the traditional sense. Along this work we have considered the class of minimal entropy configurations. These don't have to be taken as the class which exactly describes the Universe along its evolution: they simply correspond to how the Universe mostly "looks like". At any space-time volume, the weight of these configurations is not so picked around the minimum of entropy: configurations with a slightly higher entropy contribute just a little bit less to the sum 11.3. However, a small entropy difference between two configurations means that they also look mostly the same, in the sense that the mean values of observables, corresponding to the two configurations, are almost indistinguishable. In this sense, the "stabilization" of the solution increases as the volume of the phase space increases. For a large phase space (and the phase space gets larger and larger with volume) we obtain therefore a description of observables quite similar to the traditional concept of "mean values". Also the Einstein's equations, which are at the base of our derivation of the Heisenberg's Uncertainty Principle in a bounded Universe, have to be interpreted as "mean value equations". The existence of a differentiable geometrical description of space-time itself is a "mean value statement". In fact, quantization of space-time coordinates leads to the existence of a minimal length; the minimal geometric unit is therefore not the point, but a Planck-size cell. Under these conditions, concepts like "open" versus "closed" topology are deprived of any meaning, and it becomes possible to map from a representation of the Universe in which the horizon is the spheric surface of a ball with a certain radius, and the observer located at the center of the ball, to a dual picture in which the horizon is represented as the origin of the Universe. Indeed, the operation leading to this identification, at the origin of the curvature of space-time, is a kind of T-duality.

Within the class of minimal entropy configurations, a four-dimensional space-time is selected, and T-duality is softly broken; it does then make sense to use at large volumes an approximated description in terms of classical geometry. It is under this condition that we can also assign an ordering to the configurations of the Universe according to a "time" evolution. The identification of a coordinate with our concept of time is in fact related to the existence of a classical geometry, and as such this too is an average, "mean value-like" property. Owing to the fact that a minimal entropy configuration at space-time volume V (and time $\mathcal{T}$) has higher probability of projecting to configurations at volume $V + \delta V$ (time $\mathcal{T} + \delta\mathcal{T}$) than to configurations at $V - \delta V$ (time $\mathcal{T} - \delta\mathcal{T}$), the time evolution occurs in the average toward an increase of entropy, $\partial S / \partial t \geq 0$ (second law of thermodynamics).

The weakening of the concept of "solution" to the one of simple "weighted superposition" of string configurations effectively provides a non-field theoretical realization of the idea of spontaneous breaking of symmetry. At early times, the weights of very different configurations are relatively close to each other; as time goes by, and the Universe cools down, the mean value of observables becomes the more and more dominated by the contribution of minimal entropy, less symmetric configurations. In the class of the dominant configurations are broken not only supersymmetry and the space parity/time reversal symmetries, but also the invariance under space-time translations and rotations: all the symmetries that appear to be broken at the macroscopical level are broken also in the fundamental description. This is due to the embedding of entropy at a fundamental, "quantum gravity level", implied in 11.3, which describes the physics of the Universe "on shell", i.e. as it concretely is. In

this way, the macroscopical and microscopical description of the physical world are sewed together without conceptual separation.

Expression 11.3 is therefore somehow a new form of path integral; it can be shown to reduce to the ordinary one in the "classical" limit, i.e. along the solutions of minimal entropy at large times. In this limit, the exponential reduces to the ordinary Lagrangian action without potential term. There is on the other hand no need of such a term. In quantum field theory, interactions are introduced through a covariantization of the kinetic term; the potential is only needed in order to introduce masses via a Higgs mechanism. But in our scenario masses have another origin: they are somehow "boundary" terms. In a bounded space-time, the boundary terms cannot be neglected, and it turns out that mass and energy densities precisely scale as the inverse of a two-dimensional surface, such as the horizon is, rather than the inverse of a three-dimensional volume. At present time, they are therefore of the order of the inverse square of the age of the Universe, in agreement with what experimental observations suggest. From this point of view, we could roughly say that the Higgs potential, introduced in an action defined on an infinitely extended space-time, is somehow an effective parametrization of boundary effects due to compactness of space-time.

A correct treatment of mass terms and the derivation of precise predictions about the physical world require on the other hand different tools than those of field theory. The physical vacuum, intended in the above sense of the dominant configuration, turns out to be strongly coupled, i.e. non-perturbative with respect to the usual asymptotic states. While ordinary quantum field theory, and the ordinary path integral, can be approached with a perturbative expansion built around free, plane wave states, the asymptotic states of 11.3 are entire "asymptotic configurations", full, non-perturbative string vacua. The problem is therefore to "solve" for the physical content of these "asymptotic string vacua". For this we had to use non-perturbative string techniques, and, in order to investigate the low-energy world, introduce approaches different from the usual one, which is based on an expansion over Feynman diagrams, even in the form of string diagrams. This is in fact of no help to our purpose, because it assumes a perturbative expansion corresponding to a linearized, "logarithmic" representation of the vacuum. In order to investigate the non-perturbative configuration we have instead used a method more closely related to the functional 11.3, based on the analysis of entropy in the renormalization processes. In this way, we could obtain all mass and coupling values of the sub-Planckian world, i.e. what manifests to us in terms of elementary particles and fields, with a reasonable degree of accuracy. The resulting spectrum turns out to be the known one of the Standard Model (though with massive neutrinos), but without Higgs field. There is also no low-energy supersymmetry and no "new physics" before the unification of couplings, which takes place at the Planck scale. Nevertheless, everything is consistent, because it arises and is explained within a non field-theoretical scheme. Indeed, from the point of view of this scenario, the reason why in field theory low-energy supersymmetry improves the field theory scaling of masses and couplings has to be put in relation to the fact that the linearization introduced in such a representation of the low-energy world is precisely built around a vanishing value of the supersymmetry breaking parameter. The unification properties of a deeply non-supersymmetric world are then mapped to a supersymmetric representation of the degrees of freedom.

In our framework, all masses and couplings turn out to be functions of the age of the Universe. At present time, their values agree, within the approximations introduced in the computing procedure, with what experimentally measured (in the case of neutrinos, our computation remains a not yet tested prediction). Besides the phenomenology of elementary particles, we have also investigated the implications of the scenario for cosmology. In its whole, the Universe turns out to behave as a black hole, with a total energy related to the radius, i.e. the age, according to the Schwarzschild relation. Inside the black hole, namely, within the horizon, the energy density is below the Schwarzschild threshold, and the Universe appears as a "normal" object; the $1/R^2$ scaling of energy and mass densities is precisely the scaling expected in such a black hole. The dominant configurations of 11.3 predict energy and matter densities precisely in the amount necessary to give space-time the curvature of a 3-sphere, as expected also on the base of considerations over light paths and the equivalence horizon/origin of space-time. Interestingly, this "classical" curvature is given by the sum of all energy densities, namely the cosmological one and those corresponding to matter and radiation. Therefore, a curvature that from a certain point of view could be justified as entirely due to the properties of light, from another point of view is seen to be produced also by matter and a "ground" source of gravity (the cosmological term). This is the result of an underlying symmetry of the dominant string configurations, which exchanges radiation, matter and gravity. If General Relativity allowed to understand how matter "generates" gravity, and Quantum Field Theory how light-to-matter, matter-to-light transitions occur, this non-perturbative string scenario closes the circle, allowing to see the duality of all these manifestations of the physical world, and therefore also how gravity converts into matter or radiation.

Apart from the absence of a Higgs field, other phenomenological implications that depart from what commonly expected are to be found in the expansion of the Universe, whose acceleration is for us only apparent; in the so-called "time variation of the fine structure constant α"; and in the non-existence of dark matter. In this scenario, for each of these phenomena the explanation relies in the particular evolution of mass scales and couplings, as functions of the age of the Universe. In any case, our predictions are compatible with the experimental observations. Indeed, one of the things that characterize this scenario and distinguish it from many other ones, is its high predictive power, due to the fact that there are no free parameters other than the age of the Universe itself. Any measurable quantity is determined, in terms of the fundamental constants (the speed of light, the Planck constant and the Planck mass), as a function of the age of the Universe. It is therefore not a trivial fact that, besides its success in producing values of observables (masses, couplings, CP violation parameters etc...) correct within the degree of approximation we used, this scenario also survives to the stringent tests coming from cosmology, which is well known for imposing severe constraints on model building.

The assumption at the base of our analysis of the dominant configurations of 11.3 has been that only critical superstring theory, with non-perturbatively stable, tachyon-free vacua, had to be considered. It is nevertheless an extremely actual and open question, whether tachyons can show out at some energy scale. Related to this is also the question about why a Minkowskian space should be selected, instead of a Euclidean one. Indeed, Minkowskian

space, i.e. the assignment of a signature $(-1, +1)$ for the rotation between two extended coordinates, after which a time-like coordinate is unambiguously selected, could be viewed as the result of the breaking of a symmetry: the $(-1, +1)$ space is not anymore symmetric in the two coordinates. In our scenario, the breaking of a symmetry reduces entropy, thereby leading to a configuration favoured with respect to the more symmetric one. Although we don't have at disposal tools refined enough to rigorously prove the statement, there are good reasons to believe that Euclidean vacua correspond to more entropic configurations than Minkowskian ones, and therefore their contribution to 11.3 is suppressed by a lower weight [74]. For similar reasons, also the supersymmetric string should be more favoured than the bosonic one, because a phase space built on spinorial representations can be differentiated more than one built on integer spin representations (a space built on spinors contains also bosons, but not vice-versa). Entropy can be reduced more on a spinorial space than on a just bosonic one. Analogous arguments could also select the critical number of dimensions with respect to non-critical ones: moving away from the critical dimension entails switching on additional degrees of freedom. Strings in non-critical dimension can in fact be parametrized as deformations of strings in critical dimension, driven by additional fields (Liouville fields). The process of entropy minimization would then probably require to freeze these degrees of freedom.

Analogous arguments can be applied also to light: if we think to embed General Relativity (and String Theory) into a wider class of theories, describing vacua in which the speed of light is in general no more a universal constant, c can be viewed as the expectation value of a "field": $c \equiv < c(X) >$, where X stays for a collection of coordinates in the higher dimensional space of these theories. Freezing c to a constant can then be seen as the consequence of a process of reduction of the degrees of freedom, required by entropy minimization. The class of vacua with c constant results therefore favoured, i.e. weights more in 11.3, than the vacua with variable c. The selection of the actual value of c among all the possible ones is then just a matter of fixing the units of measure. But there is more: the very fact of beeing the speed of light *finite* can be seen as a consequence of the minimization of entropy. If we include also infinite volume configurations, together with finite volume ones, those with finite volume dominate because of their lower entropy, due to the smaller number of combinatorics, as it can be easily seen by considering that in our ordering infinite volume is the limit toward which the system tends, through steps of increasing entropy. A space-time of finite volume can only exist if the speed of light is finite. Therefore, saying that the configurations with finite volume dominate over those with infinite volume implies that a finite speed of light is the situation which is realized with the highest weight. According to 11.3, the Universe is then basically the one arising from a finite speed of light, something that "closes the circle" leading us back to the statement that was at the beginning of the entire investigation. We started in section 2 with a discussion of the consequences of the finiteness of the speed of light, to conclude for the quantum nature of physics, and arrive to 11.3. Now we understand that perhaps we have at hand something deeper. The functional 11.3 seems to be a really

[74] In Ref. [1] we approach the problem from a different perspective: it turns out that in the phase space of all possible configurations, the dominant one is the one in which there exists a field, the photon, that propagates with the same speed the Universe itself evolves, namely with the fundamental speed at which entropy changes, thereby giving to the observer the perception of (time) evolution.

fundamental and general expression, which accounts also for the selection of the "true", consistent string vacua, out of all the possible string constructions, including bosonic and tachyonic ones, and even beyond the basic assumption of the theory of Relativity, namely the universality of the speed of light. Configurations with constant speed of light, and, among these, critical superstring theory, would then be selected because favoured by the fact of containing in their class configurations of lower entropy. On the other hand, the fact that 11.3 doesn't lead to exact solutions in the classical sense, but simply decides what is the amount of contribution of any configuration to the appearance of the Universe, would then imply that, although suppressed, tachyonic and non-critical string aspects are not completely absent from our Universe.

Despite the progress made with respect to Ref. [2], even our present degree of knowledge is far from being complete. The present work represents only an improvement in a research project whose first steps go back to Refs. [16, 11], when examples of non-perturbative string-string duality between constructions in four space-time (infinitely extended) dimensions and non-compact orbifolds in lower dimensions, first suggested that the natural environment, the target space-time, in which to define string theory, instead of being the one of infinite volume, to be then appropriately compactified, could have better been a generally compact one. In this perspective, infinitely extended space-time should be regarded as a special limit, rather than the "normal" starting condition. If string theory had to be "the" theory of the Universe, then this was a strong indication that space-time could be compact.

We hope to improve in the future the methods of approaching the "solution" of the functional 2.31, the key point of this work.

Appendix

A Conversion units for the age of the Universe

We give here some conversion factors from time units to Planck mass units.

$$1 \text{ year (yr)} \;=\; 3,1536 \,\times\, 10^{7} \text{ s}$$

In order to convert this value to eV units we divide by $\hbar = 6,582122 \,\times\, 10^{-22}$ MeV s. We obtain:

$$1 \text{ yr} \;=\; 4,791160054 \,\times\, 10^{28} \text{ MeV}^{-1}$$

Considering that the Planck mass $\mathrm{M_P} \;=\; 1,2 \,\times\, 10^{19}$ GeV, we have also the relation:

$$1 \text{ yr} \;=\; 3,992633379 \,\times\, 10^{50}\, \mathrm{M_P^{-1}} \,.$$

The age of the Universe $\mathcal{T}$, estimated to be around 11,5 to 14 billion years, reads therefore:

$$\mathcal{T} \,\approx\, \left\{ \begin{array}{l} 4,59152839 \\ 5,58968673 \end{array} \right. \times\, 10^{60}\; \mathrm{M_P^{-1}}$$

If instead we take the neutron mass as the most precise way of deriving the age of the Universe, from expression 6.17 and the present-day measured neutron mass, we obtain:

$$\mathcal{T} \;\approx\; 5.038816199 \,\times\, 10^{60}\; \mathrm{M_P^{-1}} \quad (=\; 12,6202827 \times 10^{9} \text{ yr})$$

B The type II dual of the $\mathcal{N}_4 = 1$ orbifold

We discuss here the type II dual construction of the $\mathcal{N}_4 = 1$ orbifold vacuum of section 3.2.1. On the heterotic side, this appears as a supersymmetric construction. We claimed that $\mathcal{N}_4 = 1$ supersymmetry exists only perturbatively, but when the full, non-perturbative construction is considered, one sees that this symmetry is broken. From the heterotic point of view, the breaking is non-perturbative, being produced by a "twist" along the coupling-coordinate around which the perturbative expansion is built. The only signal of the supersymmetry breaking is then indirectly provided by the way couplings of non-perturbative matter and gauge sectors (parametrized by perturbative fields of the heterotic string) enter in the expressions of threshold corrections of effective couplings. Namely, with the "wrong" power, as if these couplings were "inverted", from a < 1 to a > 1 value. Indeed, these couplings are parametrized by moduli only at the $\mathcal{N}_4 = 2$ level (see Ref. [11]). When the perturbative supersymmetry is reduced to $\mathcal{N}_4 = 1$, these fields are twisted. This however only means that the expectation value is not anymore a running parameter, but is fixed. We can nevertheless trace the fate of the couplings by investigating the so-called "$\mathcal{N} = 2$" sectors.

In order to follow the operation of supersymmetry breaking from the the type II side, let's first consider the starting point, the $\mathcal{N}_4 = 2$ construction. In order to make easier the investigation of the projections, it is convenient to express the degrees of freedom in terms of free fermions (Ref. [50]). In the case of type II strings, these constructions have been extensively analysed in Ref. [27]. Indeed, the cases we are referring to are embedded in a infinitely extended space-time, a situation deeply different from the one considered in this paper, where space-time is compact. However, as we have seen, in practice this reflects in a different interpretation of the results (e.g. the fact that densities become global quantities), whereas from a technical point of view the usual analysis carries over from a scenario to the other one with minor, obvious changes (the substitution of a continuum of modes along the space-time coordinates with a discrete lattice of momenta/energies). For simplicity, we use therefore here the same notation for the string modes as in the cited works. The set of all fermions is therefore:

$$F = \left\{ \begin{array}{c} \psi^L_\mu, \chi^L_I, y^L_I, \omega^L_I \\ \psi^R_\mu, \chi^R_I, y^R_I, \omega^R_I \end{array} \right\}, \quad (\mu = 1, 2;\ I = 1, ..., 6), \tag{2.1}$$

where $\psi^{L,R}_\mu$ indicate the left and right moving fermion degrees of freedom along the transverse space-time coordinates, while $\chi^{L,R}_I$ those along the internal coordinates. $y^{L,R}_I \omega^{L,R}_I$ correspond instead to the internal fermionized bosons. The basic sets of boundary conditions are S and $\bar{S}$, which contain only eight left- or right-moving fermions, and distinguish the boundary conditions of the left- and right- moving world-sheet superpartners:

$$S = \left\{\psi^L_\mu, \chi^L_1, \ldots, \chi^L_6\right\}, \qquad \bar{S} = \left\{\psi^R_\mu, \chi^R_1, \ldots, \chi^R_6\right\}. \tag{2.2}$$

In order to obtain a $Z_2 \times Z_2$ symmetric orbifold, we need then the two sets b_1 and b_2:

$$b_1 = \left\{ \begin{array}{c} \psi^L_\mu, \chi^L_{1,2}, y^L_{3,\ldots,6} \\ \psi^R_\mu, \chi^R_{1,2}, y^R_{3,\ldots,6} \end{array} \right\}, \tag{2.3}$$

$$b_2 = \left\{ \begin{array}{c} \psi^L_\mu, \chi^L_{3,4}, y^L_{1,2}, y^L_{5,6} \\ \psi^R_\mu, \chi^R_{3,4}, y^R_{1,2}, y^R_{5,6} \end{array} \right\}. \tag{2.4}$$

These sets assign Z_2 boundary conditions and break the $\mathcal{N}_4 = 8$ supersymmetry to $\mathcal{N}_4 = 2$. The lowest entropy configuration is then obtained by further partial shifting of some states of the twisted sectors. We will not consider these further operations: they commute with the projection we want to consider in the following, namely the one that leads to the breaking of supersymmetry; considering them complicates the construction without altering the conclusions. As discussed in Ref. [27] and [11], depending on the relative phase of the projections introduced by b_1 and b_2, we obtain two mirror configurations which, according to [11], are two slices of the same model: in one we see only the vector multiplets, in the other only the hypermultiplets, of the same $U(16)$ model.

We want now to introduce another projection, dual to the $Z_2^{(2)}$. In order to understand what we have to expect from this further operation, we must take into account that 1) in order to preserve the pattern of duality with the heterotic and type I string established at the $\mathcal{N}_4 = 2$ level, i.e. the identification of the geometric moduli of the type II space with

those of the heterotic space and the type I coupling moduli, also this third projection must act symmetrically on left and right movers; 2) it must twist all these moduli. On the other hand, we cannot pretend to see the extended space-time represented in a similar way in both the heterotic/type I and the type II dual: a further symmetric, independent twist on the type II space must necessarily act also on the coordinates with index "μ". This means that, in order to see the action of the heterotic $Z_2^{N=2 \to 1}$ projection, on the type II side we must trade the space-time coordinates for internal ones. From the type II point of view the heterotic space-time will therefore be entirely non-perturbative, and the type II construction will look perturbatively compactified to two dimensions. Being two coordinates hidden in the light-cone gauge, we see therefore no transverse non-compact coordinates. As we discussed in section 3.2.3, representing the "11-th coordinate" of string theory in orbifolds entails its linear realization through an embedding in a two-dimensional toroidal space. The space gives therefore the fake impression of being "12-dimensional". This is however an artifact of the perturbative representation.

Compactifying the "μ" indices implies that we can now fermionize the bosons also along these coordinates. The boson degrees of freedom ∂X_μ and $\bar{\partial} X_\mu$ will be now represented as $y^L_\mu \omega^L_\mu$ and $y^R_\mu \omega^R_\mu$. (By the way, we remark that in the scenario discussed in this work, all string coordinates are always compactified. Therefore, in principle fermionization of the space-time degrees of freedom is always possible. On the other hand, when considering explicit string constructions, perturbation is always possible only around a decompactified coordinate, that works as the vanishing coupling around which to perturb. The very fact of writing a perturbative representation of a string vacuum implies the assumption that a certain limiting procedure toward a non-fermionizable point of some coordinates has been taken.)

From this two-dimensional point of view, the $\mathcal{N}_4 = 2$ type II construction contains only scalar fields: the space-time is non-perturbative, and therefore so are all indices (vector, spinor and tensor) running along space-time coordinates. The type II construction is therefore blind to the distinction between gauge and matter, whose degrees of freedom have a space-time vector or spinor index, and an internal, scalar index: only this last index is visible on the type II, and these states appear all as scalars. There is no trace of the graviton, because it bears only space-time indices. Moreover, the fields T^i and U^i, $i = 1, 2, 3$, usually appearing in one-loop expressions of threshold corrections, don't correspond now to geometric moduli of two-tori. Indeed, for any twist what remains untwisted is a four-torus. In practice, we have added a two-torus. However, as we discussed, this is an artifact of the linearization of the space; there is indeed no twelve-dimensional theory, and the appearance of the two-torus is due to an "over-dimensional" representation of a curved space with just one more coordinate, the one that served as the coupling on which to expand in the four-dimensional vacuum. The 12-th coordinate is instead a curvature. There is no surprise that, in this representation, the former moduli T^i and U^i are now multiplied by what was the coupling of the theory: its dependence was simply "frozen" by construction. For what matters duality with the heterotic construction, nothing changes, because the value of these fields was not fixed. We can recover a description in terms of moduli of two-tori by introducing independent boundary conditions for the "complex planes" (1,2), (3,4), (5,6),

(7,8) (see Ref. [27] for a detailed discussion of these sets). This allows to disentangle the two-torus moduli, by factorizing the space in four two-tori. On the type II side we see then that, besides the T^i and U^i, $i = 1, 2, 3$, we have now one more field, corresponding to what was the (hidden) coupling of the four-dimensional construction. It misleadingly appears as a pair of torus moduli, T^4, U^4, respectively corresponding to the volume form and the complex structure. Owing to the symmetry of the construction under exchange of the three tori with the fourth one, a $T^4 \leftrightarrow U^4$ reflection exchanges the two $\mathcal{N} = 4$ mirror constructions (the one with only vectors with the one with only hyper multiplets). It is worth to consider more in detail this property. The "fourth torus" volume form is the product of two radii, that we call R_{11} and R_{12} for obvious reasons. The moduli T^4 and U^4 are related to these radii by: $\operatorname{Im} T^4 = R_{11}R_{12}$, $\operatorname{Im} U^4 = R_{11}/R_{12}$. As we said, one of the two radii is indeed not a real further coordinate, but a curvature. When seen from the "four dimensional point of view", an inversion of this radius corresponds to an inversion of the full string coupling. Therefore, the $T^4 \leftrightarrow U^4$ mirror exchange that relates the two constructions is an "S-duality" of the "normal representation" of the type II vacuum.

We already discussed in Ref. [11] how the heterotic construction, containing both vector and hyper multiplets, corresponds to a slice, built around a corner of the moduli space, of the "union" of both the type II mirror models. From this point of view it is therefore "self-mirror". Here we understand that this mirror symmetry is indeed a strong-weak coupling duality of the type II string, an operation which is perturbative on the heterotic dual [75]. For the rest, it is important to observe that, although we cannot explicitly verify it on the base of the carried space-time indices, all hidden, the identification of the degrees of freedom allows anyway to see the S and $\bar{S}$ as the generators of space-time supersymmetry. This time they are to be intended as a representation of the "internal part" of the supersymmetry sets.

From the above considerations, we conclude that, on the type II side, the new projection, corresponding to the step $\mathcal{N}_4 = 2 \rightarrow \mathcal{N}_4 = 1$, must be represented by a set b_3 given, up to a permutation of the three complex planes corresponding to the indices $I = 1, \ldots, 6$, by:

$$b_3 = \left\{ \begin{array}{l} \chi_{3,\ldots,6}, y^L_\mu, y^L_{1,2} \\ \chi^R_{3,\ldots,6}, y^R_\mu, y^R_{1,2} \end{array} \right\} . \tag{2.5}$$

The condition 2) of above tells us however that, differently from the case of b_1 and b_2, the "GSO phase" of this set must be [76]:

$$\delta_{b_3} = -1 \, , \tag{2.6}$$

(we recall that $\delta_{b_1} = \delta_{b_2} = 1$ and $\delta_S = \delta_{\bar{S}} = -1$). This condition projects out all the states of the type $\phi^L \otimes \phi^R$, for whatever indices and $\phi \in \{\psi, \chi, y, \omega\}$, i.e. all the states of the untwisted sector. The moduli "T" and "U" are now "twisted", and the only massless states come from the twisted sectors. The projection coefficients of the fermionic construction are

[75] On the heterotic side, matter and gauge sectors are exchanged by an exchange of the twisted and the untwisted sectors. This corresponds to an inversion of the world-sheet parameter τ: $\tau \rightarrow -1/\tau$. This parameter is integrated out, and it never appears explicitly in the effective theory. On the other hand, we have seen that the world-sheet coordinates are roughly "identified" with the two longitudinal coordinates of the light-cone gauge. Any trace of the moduli of this symmetry is therefore hidden by the gauge fixing.

[76] We refer the reader to [50] for an explanation of this coefficient and its role.

given in the following table:

	F	S	$\bar{S}$	b_1	b_2	b_3
F	1	-1	-1	1	1	1
S	-1	1	1	-1	-1	-1
$\bar{S}$	-1	1	1	-1	-1	-1
b_1	1	1	1	1	1	1
b_2	1	1	1	1	1	1
b_3	1	1	1	1	1	1

(2.7)

together with the conditions: $\delta_S = \delta_{\bar{S}} = \delta_{b_3} = -1$, $\delta_\phi = \delta_{b_1} = \delta_{b_2} = 1$. Observe that, with this choice, b_3, although a type II symmetric twist as b_1 and b_2, projects the states with the same phase as a heterotic Z_2 orbifold projection, as we precisely wanted. Notice also that, differently from how it appears on the heterotic side, the projection introduced by b_3 is not exactly symmetrical to the one introduced by b_2. For instance, it seems that it would project out all the T and U fields even when acting alone, i.e. before the introduction of b_2. This impression is however misleading, in that it neglects that, as we have seen, from the point of view of this two-dimensional compactification, these fields are no more moduli of a torus, but have a more complicate expression as functions also of the former coupling coordinate, here "embedded" in the further, fourth torus. And indeed, if we want to introduce the "planes" as in Ref. [27, 11] in order to lower the rank of the twisted sectors, the sets which introduce separate boundary conditions for the coordinates must be defined in order to include more than one bosonic coordinate. Namely, they must contain also the "coupling plane". In the $\mathcal{N}_4 = 2$ model constructed with just $\{b_3, b_1\}$ (or $\{b_3, b_2\}$) the moduli T^i, U^i are no more built from the states:

$$\delta_{ij}\, x_i \bar{x}_j |0> \tag{2.8}$$

but as combinations of states of the type:

$$x_i \bar{x}_j |0> \qquad i \neq j\,,\ \{i,j\} \in (\{3,4\},\, \{5,6\}, \{7,8\}) \ \cup \{11,12\}\,. \tag{2.9}$$

The partition function of this orbifold is given by the integral over the modular parameter τ, with modular-invariant measure $(\,\mathrm{Im}\ \tau)^{-2} d\tau d\bar{\tau}$, of:

$$Z^{\mathrm{string}} = \left(\frac{1}{2}\right)^3 \sum_{(H_1,G_1,H_2,G_2,H_3,G_3)} Z^F_{\mathrm{L}}\ Z^F_{\mathrm{R}} \sum_{(\gamma,\delta)} Z_{8,8}\begin{bmatrix}\gamma\\ \delta\end{bmatrix}\,, \tag{2.10}$$

where $Z^F_{L,R}$ contain the contribution of the world-sheet fields $\psi^{L,R}_\mu$, $\chi^{L,R}_a$ (the sets S and $\bar{S}$); $Z_{8,8}$ substitutes what in four dimensional constructions is $Z_{6,6}$, the $c = (6,6)$ internal space. Now this space spans all bosonic degrees of freedom and has $c = (8,8)$, corresponding to the fields $\omega^{L,R}_I$, $y^{L,R}_I$, $I = 1, \ldots, 8$. Notice that we don't have now the factor $1/(\,\mathrm{Im}\,\tau|\eta(\tau)|^4)$, the contribution of the space-time transverse bosonic degrees of freedom, now accounted in

$Z_{8,8}$. We have:

$$Z_{\mathrm{L}}^{F} = \frac{1}{2}\sum_{(a,b)} \frac{\mathrm{e}^{i\pi\varphi_{\mathrm{L}}\left(a,b,\vec{H},\vec{G}\right)}}{\eta^4}\vartheta\begin{bmatrix}a+H_3\\ b+G_3\end{bmatrix}\vartheta\begin{bmatrix}a+H_2-H_3\\ b+G_2-G_3\end{bmatrix}\vartheta\begin{bmatrix}a+H_1\\ b+G_1\end{bmatrix}\vartheta\begin{bmatrix}a-H_1-H_2\\ b-G_1-G_2\end{bmatrix}\;, \tag{2.11}$$

$$Z_{\mathrm{R}}^{F} = \frac{1}{2}\sum_{(\bar{a},\bar{b})} \frac{\mathrm{e}^{i\pi\varphi_{\mathrm{R}}\left(\bar{a},\bar{b},\vec{H},\vec{G}\right)}}{\bar{\eta}^4}\vartheta\begin{bmatrix}\bar{a}+H_3\\ \bar{b}+G_3\end{bmatrix}\vartheta\begin{bmatrix}\bar{a}+H_1-H_3\\ \bar{b}+G_1-G_3\end{bmatrix}\vartheta\begin{bmatrix}\bar{a}+H_2\\ \bar{b}+G_2\end{bmatrix}\vartheta\begin{bmatrix}\bar{a}-H_1-H_2\\ \bar{b}-G_1-G_2\end{bmatrix}\;, \tag{2.12}$$

with:

$$\varphi_{\mathrm{L}} = a+b+ab\;, \tag{2.13}$$
$$\varphi_{\mathrm{R}} = \bar{a}+\bar{b}+\bar{a}\bar{b}\;. \tag{2.14}$$

The contribution of the compact bosons is:

$$\begin{aligned} Z_{8,8}\begin{bmatrix}\gamma\\ \delta\end{bmatrix} &= \mathrm{e}^{i\pi(H_3+G_3+H_3G_3)} \\ &\times\ \frac{1}{|\eta|^4}\left|\vartheta\begin{bmatrix}\gamma\\ \delta\end{bmatrix}\vartheta\begin{bmatrix}\gamma+H_3\\ \delta+G_3\end{bmatrix}\right|^2 \\ &\times\ \frac{1}{|\eta|^4}\left|\vartheta\begin{bmatrix}\gamma\\ \delta\end{bmatrix}\vartheta\begin{bmatrix}\gamma+H_2+H_3\\ \delta+G_2+G_3\end{bmatrix}\right|^2 \\ &\times\ \frac{1}{|\eta|^4}\left|\vartheta\begin{bmatrix}\gamma\\ \delta\end{bmatrix}\vartheta\begin{bmatrix}\gamma+H_1\\ \delta+G_1\end{bmatrix}\right|^2 \\ &\times\ \frac{1}{|\eta|^4}\left|\vartheta\begin{bmatrix}\gamma\\ \delta\end{bmatrix}\vartheta\begin{bmatrix}\gamma+H_1+H_2\\ \delta+G_1+G_2\end{bmatrix}\right|^2\;. \end{aligned} \tag{2.15}$$

The pairs (a,b) and $(\bar{a},\bar{b})$ specify the boundary conditions, in the directions $\mathbf{1}$ and τ of the world-sheet torus, of the sets S and $\bar{S}$, while (γ,δ) refer to the set of all fermionized bosons; (H_1,G_1), (H_2,G_2) and (H_3,G_3) refer to the sets b_1, b_2 and b_3. Notice the presence of the phase $\mathrm{e}^{i\pi(H_3+G_3+H_3G_3)}$, corresponding to the choice $\delta_{b_3}=-1$.

In this model there are nine massless sectors, corresponding to the previous b_1, b_2, b_1b_2, the new ones, b_3, b_3b_1, Fb_3b_2, $b_3b_1b_2$, $S\bar{S}b_3b_2$, and the $S\bar{S}$ sector. Only three sectors have a perturbative dual on the heterotic side, and correspond to a tern generated by a pair of intersecting projections. Here $b_3\cap b_1\neq\emptyset$ and $b_2\cap b_1\neq\emptyset$, while $b_3\cap b_2=\emptyset$, therefore the pair is either $\{b_3,b_1\}$ or $\{b_2,b_1\}$. On the sets generated by one of these pairs, the third independent projection doesn't impose any further constraint. The third projection is already "built-in" by construction in the heterotic string, which starts with half the maximal supersymmetry of the type II string. Therefore, apart from the supersymmetry reduction, from the heterotic point of view the further projection triplicates the structure of the $\mathcal{N}_4=2$ model. However, on the type II side, where we have access to all the sectors, we can see that some of the sectors hidden for the heterotic string are not supersymmetric: owing to the δ_{b_3} GSO torsion,

the $S\bar{S}$ states are here supersymmetric to nothing, and the same is true for the states of the Fb_3b_2 and $S\bar{S}b_3b_2$ sectors: their superpartners are massive. This is a representation in terms of free fermions of what more generally is a mass shift (see Ref. [27] for a discussion of the translation of the fermionic language in terms of orbifold operations).

With different choices of the relative GSO projections of one sector to the other one, the coefficients $(b_3|b_j)$ in table 2.7, we obtain mirror configurations in which supersymmetry is broken in a different way: a negative projection of b_3 to b_1 and b_2 implies that all the twisted sectors are projected out. Some of them, not as a consequence of a shift, but due to incompatibility of the selected chiralities of the spinors of the twisted sectors. It seems therefore that the model is empty unless the S and $\bar{S}$ projections are removed from the definition of the basis: only the pure Ramond-Ramond sector survives (the projections $(b_3|S)$ and $(b_3|\bar{S})$ remain unchanged). These mirror models seem to exist only at a "delta-function" point in the string moduli space.

C The supersymmetry-breaking scale

The string vacuum whose type II dual has been discussed in appendix B shows that, once the string space attains the maximal amount of twisted coordinates, supersymmetry is broken and the space is necessarily curved. In section 3.2.1 we pointed out that, in a compact space, supersymmetry is always broken, as a consequence of the missing invariance under space-time translations, which are part of the super-Poincaré group ($\{Q, \bar{Q}\} \sim P$). Indeed, minimization of entropy requires that all coupling moduli, and in particular the heterotic dilaton field, are twisted, and frozen at the Planck scale. All this means that the heterotic construction discussed in sections 3.2, 3.2.1 doesn't correspond to a true perturbation around a decompactified coordinate, but is a kind of non-compact orbifold, in which, in order to be able to build the states around a small/vanishing value of a coordinate, we artificially neglect its being twisted, and proceed as if along this direction we would not have fixed points. In some sense, this is also what is done in Ref. [43, 44]. However, in that case, as long as one is not interested in further curling of the string space, and remains at the level of maximal supersymmetry, the game may appear basically innocuous: it works because there are many other de-compactifiable coordinates. Problems arise when proceeding to further twisting, as we observed in Ref [11].

The result is that there are "hidden" sectors, whose origin has to be traced in the original, neglected T-duality of the theory, which are non-perturbative, and where all the breaking of supersymmetry is relegated. All this is the misleading consequence of an artificial, "illegal" flattening of an intrinsically curved space. The coordinate, or better, the curvature, which is related to the size of the flattened coordinate, works as "order parameter" for the breaking of supersymmetry. Some aspects of this phenomenon are precisely responsible for what we observe on the type II side.

As we have seen in appendix B, on the type II side we have two mirror situations, in which the breaking of supersymmetry manifests itself in a different way. However, both of them can be related to the same mechanism; they must be seen as two aspects of the same phenomenon. The key point is that non-freely acting projections can be viewed as

obtained at the corner of the moduli space of freely acting constructions. In this case, at a generic point in the orbifold moduli space, projected states receive a non-vanishing mass as a consequence of a coordinate shift associated to the orbifold twist. In the decompactification limit of this coordinate, masses become infinite and the projected states disappear from the spectrum, as they usually do in ordinary non-freely acting orbifolds. This allows us to get an idea of the scale at which supersymmetry is broken: the supersymmetric partners are lifted by a shift picked along an internal coordinate X, which is also twisted. In order to describe the GSO projection process in terms of freely acting shifts, we must look for a dual configuration in which X, that we know to be fixed by minimization of entropy at a value around the Planck scale, reaches this value as a limit at the corner of the moduli space. The map between the two descriptions, $X \to R(X)$, must therefore be some kind of logarithm, so that:

$$X \to 1 \quad \leftrightarrow \quad R \to 0/\infty \, . \tag{3.1}$$

This implies that either $R \approx \ln X$ or $1/R \approx \ln X$. As discussed in ref. [27, 11], a change in sign of the $(b_i|b_j)$, projections corresponds to the inversion of some internal radius. The mirror constructions of appendix B correspond therefore precisely to the one or the other of these two possibilities, for some of the internal coordinates. According to the mechanism of freely acting projections, in the dual picture the mass of the projected states reads:

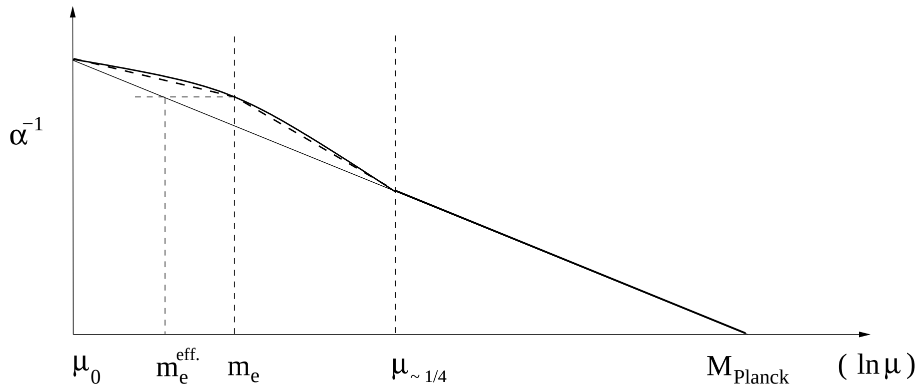
$$\tilde{m} \, (\sim \ln m) \; \sim \; R \, . \tag{3.2}$$

Pulled back to the physical picture, we obtain that, at the "twisting point", the mass is of order one in Planck units. This is the mass gap between the observed particles (and fields), with sub-Planckian mass, vanishing at first order, and their superpartners.

D Local correction to effective beta-functions

The running of the electromagnetic and weak couplings in the representation in which they are going to be compared with experimental data is logarithmic, with a slope determined by an effective beta-function coefficient. However, as discussed in section 7.4, around the scale $\sim m_e$, the volumes of the matter phase space are expanded (or, logarithmically, shifted), in such a way that for instance the electromagnetic coupling at the scale m_e (i.e. the fine structure constant) effectively corresponds to the value of the coupling *without correction* at a run-back scale, $m_e^{\text{eff.}}$. The amount of running-back in the scale of the logarithmic effective coupling is equivalent to the amount of the forward shift in the logarithmic representation of the volumes of particles in the phase space. If volumes get multiplied by a factor, their logarithm gets shifted, and so gets shifted back the scale at which the coupling in its logarithmic representation is effectively evaluated. This deviation can be considered as a perturbation of the logarithmic running, that we illustrate here. In the figure, μ_0 stays for the starting scale of the running: $\mu_0 = (1/2) \, \mathcal{T}^{-1/2}$, $\mu_{\sim 1/4}$ for the upper end scale of the matter sector, the thick solid line shows the approximate expected behaviour of the inverse coupling α^{-1}, including the correction to the shape, while the thin solid line indicates the original logarithmic behaviour. The dashed segments indicate the linear approximation of the curve we

considered in the footnote at page 148 in order to compute the effective weak coupling at the W-boson scale:



References

[1] A. Gregori, *About combinatorics, and observables,* `arXiv:0712.0471 [hep-th]`.

[2] A. Gregori, *Entropy, string theory, and our world,* `hep-th/0207195`.

[3] A. Gregori, *On the Time Dependence of Fundamental Constants,* `hep-ph/0209296`.

[4] A. Gregori, *Naturally time dependent cosmological constant in string theory,* `hep-th/0402126`.

[5] H. Culetu, *On a time dependent cosmological constant,* `hep-th/0306262`.

[6] M. R. Douglas, *Statistical analysis of the supersymmetry breaking scale,* `hep-th/0405279`.

[7] M. R. Douglas, *Basic results in vacuum statistics,* Comptes Rendus Physique **5** (2004) 965–977, `hep-th/0409207`.

[8] M. R. Douglas, *Understanding the landscape,* `hep-th/0602266`.

[9] F. Denef and M. R. Douglas, *Computational complexity of the landscape. I,* Annals Phys. **322** (2007) 1096–1142, `hep-th/0602072`.

[10] L. Susskind, *The anthropic landscape of string theory,* `hep-th/0302219`.

[11] A. Gregori, *String-String triality for d=4, Z_2 orbifolds,* JHEP **06** (2002) 041, `hep-th/0110201`.

[12] L. J. Dixon, V. Kaplunovsky, and J. Louis, *Moduli dependence of string loop corrections to gauge coupling constants,* Nucl. Phys. **B355** (1991) 649–688.

[13] E. Kiritsis and C. Kounnas, *Infrared regularization of superstring theory and the one loop calculation of coupling constants,* Nucl. Phys. **B442** (1995) 472–493, `hep-th/9501020`.

[14] E. Kiritsis, C. Kounnas, P. M. Petropoulos, and J. Rizos, *Universality properties of N = 2 and N = 1 heterotic threshold corrections,* Nucl. Phys. **B483** (1997) 141–171, `hep-th/9608034`.

[15] E. Kiritsis, C. Kounnas, P. M. Petropoulos, and J. Rizos, *Solving the decompactification problem in string theory,* Phys. Lett. **B385** (1996) 87–95, `hep-th/9606087`.

[16] A. Gregori, C. Kounnas, and P. M. Petropoulos, *Non-perturbative gravitational corrections in a class of N = 2 string duals,* Nucl. Phys. **B537** (1999) 317–343, `hep-th/9808024`.

[17] A. Gregori, C. Kounnas, and P. M. Petropoulos, *Non-perturbative triality in heterotic and type II N = 2 strings,* Nucl. Phys. **B553** (1999) 108–132, `hep-th/9901117`.

[18] E. Kiritsis, C. Kounnas, P. M. Petropoulos, and J. Rizos, *String threshold corrections in models with spontaneously broken supersymmetry,* Nucl. Phys. **B540** (1999) 87–148, `hep-th/9807067`.

[19] I. Antoniadis, E. Gava, and K. S. Narain, *Moduli corrections to gravitational couplings from string loops,* Phys. Lett. **B283** (1992) 209–212, `hep-th/9203071`.

[20] I. Antoniadis, E. Gava, and K. S. Narain, *Moduli corrections to gauge and gravitational couplings in four-dimensional superstrings,* Nucl. Phys. **B383** (1992) 93–109, `hep-th/9204030`.

[21] I. Antoniadis, H. Partouche, and T. R. Taylor, *Duality of $N = 2$ heterotic-type I compactifications in four dimensions,* Nucl. Phys. **B499** (1997) 29–44, `hep-th/9703076`.

[22] J. A. Harvey and G. Moore, *Algebras, BPS States, and Strings,* Nucl. Phys. **B463** (1996) 315–368, `hep-th/9510182`.

[23] J. A. Harvey and G. Moore, *Fivebrane instantons and R**2 couplings in $N = 4$ string theory,* Phys. Rev. **D57** (1998) 2323–2328, `hep-th/9610237`.

[24] S. Ferrara, J. A. Harvey, A. Strominger, and C. Vafa, *Second quantized mirror symmetry,* Phys. Lett. **B361** (1995) 59–65, `hep-th/9505162`.

[25] J. A. Harvey and G. Moore, *Exact gravitational threshold correction in the FHSV model,* Phys. Rev. **D57** (1998) 2329–2336, `hep-th/9611176`.

[26] A. Gregori *et al.*, *R**2 corrections and non-perturbative dualities of $N = 4$ string ground states,* Nucl. Phys. **B510** (1998) 423–476, `hep-th/9708062`.

[27] A. Gregori, C. Kounnas, and J. Rizos, *Classification of the $N = 2$, Z(2) x Z(2)-symmetric type II orbifolds and their type II asymmetric duals,* Nucl. Phys. **B549** (1999) 16–62, `hep-th/9901123`.

[28] A. Gregori and C. Kounnas, *Four-dimensional $N = 2$ superstring constructions and their (non-)perturbative duality connections,* Nucl. Phys. **B560** (1999) 135–153, `hep-th/9904151`.

[29] E. Kiritsis, N. A. Obers, and B. Pioline, *Heterotic/type II triality and instantons on K3,* JHEP **01** (2000) 029, `hep-th/0001083`.

[30] I. Antoniadis, C. Bachas, and E. Dudas, *Gauge couplings in four-dimensional type I string orbifolds,* Nucl. Phys. **B560** (1999) 93–134, `hep-th/9906039`.

[31] I. Antoniadis, C. Bachas, C. Fabre, H. Partouche, and T. R. Taylor, *Aspects of type I - type II - heterotic triality in four dimensions,* Nucl. Phys. **B489** (1997) 160–178, `hep-th/9608012`.

[32] G. Aldazabal, A. Font, L. E. Ibanez, and F. Quevedo, *Heterotic/Heterotic Duality in D=6,4,* Phys. Lett. **B380** (1996) 33–41, `hep-th/9602097`.

[33] A. Sen and C. Vafa, *Dual pairs of type II string compactification,* Nucl. Phys. **B455** (1995) 165–187, `hep-th/9508064`.

[34] E. G. Gimon and C. V. Johnson, *K3 Orientifolds,* Nucl. Phys. **B477** (1996) 715–745, `hep-th/9604129`.

[35] I. Antoniadis, J. P. Derendinger, and C. Kounnas, *Non-perturbative temperature instabilities in $N = 4$ strings,* Nucl. Phys. **B551** (1999) 41–77, `hep-th/9902032`.

[36] S. Kachru and C. Vafa, *Exact results for N=2 compactifications of heterotic strings,* Nucl. Phys. **B450** (1995) 69–89, `hep-th/9505105`.

[37] M. Berkooz and R. G. Leigh, *A $D = 4$ $N = 1$ orbifold of type I strings,* Nucl. Phys. **B483** (1997) 187–208, `hep-th/9605049`.

[38] I. Antoniadis, E. Gava, and K. S. Narain, *Moduli corrections to gauge and gravitational couplings in four-dimensional superstrings,* Nucl. Phys. **B383** (1992) 93–109, `hep-th/9204030`.

[39] A. E. Faraggi, C. Kounnas, S. E. M. Nooij, and J. Rizos, *Towards the classification of Z(2) x Z(2) fermionic models,* `hep-th/0311058`.

[40] A. E. Faraggi, C. Kounnas, S. E. M. Nooij, and J. Rizos, *Classification of the chiral Z(2) x Z(2) fermionic models in the heterotic superstring,* Nucl. Phys. **B695** (2004) 41–72, `hep-th/0403058`.

[41] A. E. Faraggi, C. Kounnas, and J. Rizos, *Chiral family classification of fermionic Z(2) x Z(2) heterotic orbifold models,* Phys. Lett. **B648** (2007) 84–89, `hep-th/0606144`.

[42] P. S. Aspinwall and D. R. Morrison, *Point-like instantons on K3 orbifolds,* Nucl. Phys. **B503** (1997) 533–564, `hep-th/9705104`.

[43] P. Horava and E. Witten, *Heterotic and type I string dynamics from eleven dimensions,* Nucl. Phys. **B460** (1996) 506–524, `hep-th/9510209`.

[44] P. Horava and E. Witten, *Eleven-Dimensional Supergravity on a Manifold with Boundary,* Nucl. Phys. **B475** (1996) 94–114, `hep-th/9603142`.

[45] R. Dijkgraaf, E. Verlinde, and H. Verlinde, *C=1 conformal fields theories on Riemann surfaces,* Comm. Math. Phys. **115** (1988) 2264.

[46] E. G. Gimon and J. Polchinski, *Consistency Conditions for Orientifolds and D-Manifolds,* Phys. Rev. **D54** (1996) 1667–1676, `hep-th/9601038`.

[47] C. Angelantonj, *Comments on open-string orbifolds with a non-vanishing B(ab),* Nucl. Phys. **B566** (2000) 126–150, `hep-th/9908064`.

[48] I. Antoniadis, E. Dudas, and A. Sagnotti, *Supersymmetry breaking, open strings and M-theory,* Nucl. Phys. **B544** (1999) 469–502, `hep-th/9807011`.

[49] I. Antoniadis, G. D'Appollonio, E. Dudas, and A. Sagnotti, *Partial breaking of supersymmetry, open strings and M- theory,* Nucl. Phys. **B553** (1999) 133–154, `hep-th/9812118`.

[50] I. Antoniadis, C. P. Bachas, and C. Kounnas, *FOUR-DIMENSIONAL SUPERSTRINGS,* Nucl. Phys. **B289** (1987) 87.

[51] S. Coleman, *Aspects of Symmetry.* Cambridge University Press, 1985.

[52] F. A. Bais, *To be or not to be? Magnetic monopoles in non-Abelian gauge theories,* `hep-th/0407197`.

[53] J. M. Bardeen, B. Carter, and S. W. Hawking, *The Four laws of black hole mechanics,* Commun. Math. Phys. **31** (1973) 161–170.

[54] J. D. Bekenstein, *Black holes and entropy,* Phys. Rev. **D7** (1973) 2333–2346.

[55] **Supernova Cosmology Project** Collaboration, S. Perlmutter *et al.*, *Measurements of Omega and Lambda from 42 High-Redshift Supernovae,* Astrophys. J. **517** (1999) 565–586, `astro-ph/9812133`.

[56] **Boomerang** Collaboration, P. de Bernardis *et al.*, *First results from the BOOMERanG experiment,* `astro-ph/0011469`.

[57] **Boomerang** Collaboration, A. Melchiorri *et al.*, *A measurement of Omega from the North American test flight of BOOMERANG,* Astrophys. J. **536** (2000) L63–L66, `astro-ph/9911445`.

[58] W.-M. Yao *et al.*, *Review of Particle Physics,* J. Phys. G: Nucl. Part. Phys. **33** (2006) 1–1232.

[59] G. 't Hooft, *ON THE QUANTUM STRUCTURE OF A BLACK HOLE,* Nucl. Phys. **B256** (1985) 727.

[60] L. Susskind, *Some speculations about black hole entropy in string theory,* `hep-th/9309145`.

[61] L. Susskind and J. Uglum, *Black hole entropy in canonical quantum gravity and superstring theory,* Phys. Rev. **D50** (1994) 2700–2711, `hep-th/9401070`.

[62] D. Kabat, *Black hole entropy and entropy of entanglement,* Nucl. Phys. **B453** (1995) 281–302, `hep-th/9503016`.

[63] A. Wyler C. R. Acad. Sci. Paris **A269** (1969) 743.

[64] J. Smith, Frank D. (Tony), *From sets to quarks: Deriving the standard model plus gravitation from simple operations on finite sets,* `hep-ph/9708379`.

[65] C. Castro, *On geometric probability, holography, Shilov boundaries and the four physical coupling constants of nature,* Prog. Phys. **2** (2005) 30–36.

[66] C. Castro, *On the coupling Constants, Geometric Probability and Complex Domains,* Prog. Phys. **2** (2006) 46–53.

[67] W. Smilga, *Spin foams, causal links and geometry-induced interactions,* `hep-th/0403137`.

[68] T. E. W. Group, *Combination of CDF and D0 Results on the Mass of the Top Quark,* `arXiv:0808.1089v1 [hep-ex]`.

[69] A. A. Penzias and R. W. Wilson, *A Measurement of Excess Antenna Temperature at 4080 Mc/s,* Astrophysical Journal **142** (1965) 414.

[70] J. C. Mather, D. J. Fixsen, R. A. Shafer, C. Mosier, and D. T. Wilkinson, *Calibrator Design for the COBE Far Infrared Absolute Spectrophotometer (FIRAS),* Astrophys. J. **512** (1999) 511–520, `astro-ph/9810373`.

[71] G. F. Smoot *et al.*, *Structure in the COBE differential microwave radiometer first year maps,* Astrophys. J. **396** (1992) L1–L5.

[72] D. Clowe *et al.*, *A direct empirical proof of the existence of dark matter,* `astro-ph/0608407`.

[73] W. K. Ford-Jr and V. Rubin, *Rotation of the Andromeda Nebula from a Spectroscopic Survey of Emission Regions,* Astrophysical Journal **159** (1970) 379.

[74] W. K. Ford-Jr, V. Rubin, and N. Thonnard, *Rotational Properties of 21 Sc Galaxies with a Large Range of Luminosities and Radii from NGC 4605 (R=4kpc) to UGC 2885 (R=122kpc),* Astrophysical Journal **238** (1980) 471.

[75] M. T. Murphy *et al.*, *Possible evidence for a variable fine structure constant from QSO absorption lines: motivations, analysis and results,* Mon. Not. Roy. Astron. Soc. **327** (2001) 1208, `astro-ph/0012419`.

[76] J. K. Webb *et al.*, *Further Evidence for Cosmological Evolution of the Fine Structure Constant,* Phys. Rev. Lett. **87** (2001) 091301, `astro-ph/0012539`.

[77] V. A. Dzuba, V. V. Flambaum, and J. K. Webb, *Calculations of the relativistic effects in many-electron atoms and space-time variation of fundamental constants,* Phys. Rev. **A59** (1999) 230–237, `physics/9808021`.

[78] X. Calmet and H. Fritzsch, *Symmetry breaking and time variation of gauge couplings,* Phys. Lett. **B540** (2002) 173–178, `hep-ph/0204258`.

[79] X. Calmet and H. Fritzsch, *The cosmological evolution of the nucleon mass and the electroweak coupling constants,* Eur. Phys. J. **C24** (2002) 639–642, `hep-ph/0112110`.

[80] V. V. Flambaum and E. V. Shuryak, *Limits on cosmological variation of strong interaction and quark masses from big bang nucleosynthesis, cosmic, laboratory and Oklo data,* Phys. Rev. **D65** (2002) 103503, `hep-ph/0201303`.

[81] T. Damour and F. Dyson, *The Oklo bound on the time variation of the fine-structure constant revisited,* Nucl. Phys. **B480** (1996) 37–54, `hep-ph/9606486`.